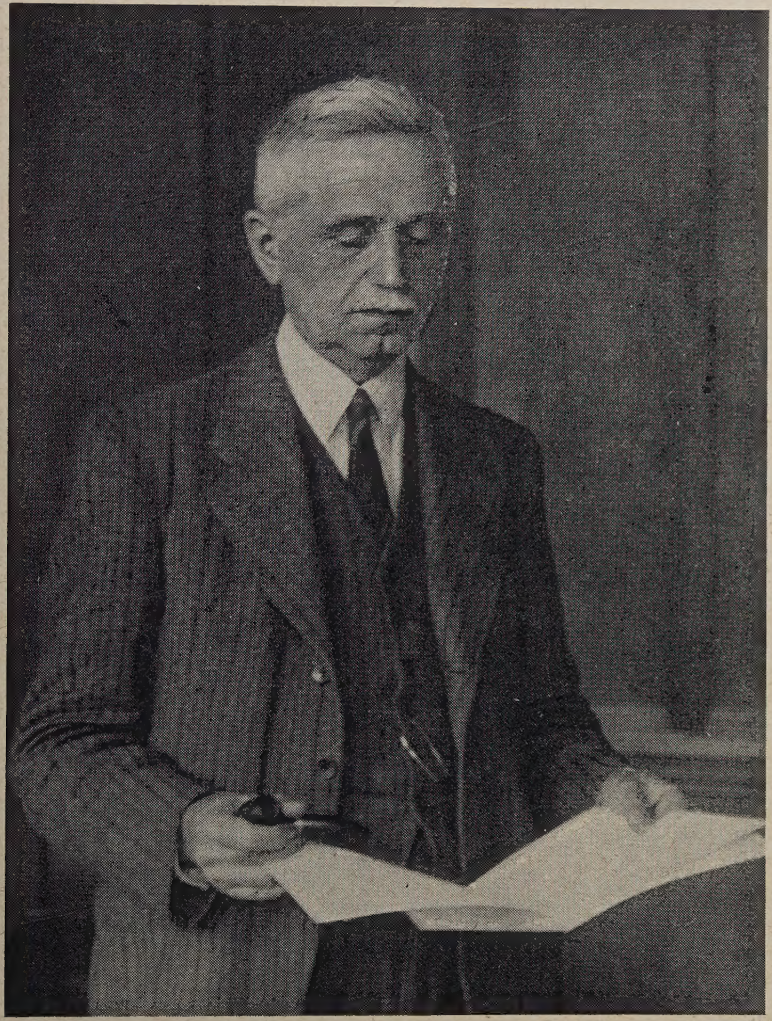
VOL. 9 No. 1, p. 1-32 1947

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, HOLLAND



On the 20th October 1946 Dr. Ekko Oosterhuis celebrated his 60th birthday. It was in that year also 35 years ago that he graduated in mechanics and physics at Groningen. After having worked for a number of years in the Kamerlingh Onnes Laboratory at Leyden, in 1914 Dr. Oosterhuis entered the service of N.V. Philips Gloeilampenfabrieken at Eindhoven as a physicist.

The editors of Philips Technical Review have pleasure in placing these facts on record, because Dr. Oosterhuis has held a very special position among us ever since this journal was first published in 1936. Thanks to his critical mind and allround knowledge he is always able to reduce the problems inherent in a subject to their root and to deal with them in a most comprehensible manner. If this journal has found so many friends among its readers this is undoubtedly due in no small degree to the efforts of Dr. Oosterhuis. It is to be hoped that we may continue to benefit from his judgment and energy for many years to come.

COLOUR STIMULUS AND COLOUR SENSATION

by P. J. BOUMA and A. A. KRUITHOF.

535.65:535.733

The difference between the two conceptions "colour sensation" and "colour stimulus" is explained. Colour sensation, besides being dependent upon the spectral distribution of the light striking our eye from a coloured object, is strongly influenced by various kinds of secondary circumstances, such as the colour of the surroundings of the object observed and the condition of our eyes. Characteristic features of a colour sensation are its hue, saturation and brightness. A colour stimulus is the subject of trichromatic measurements, by methods which, thanks to standardisation, practically exclude such secondary circumstances. The results of the measurement of a colour stimulus may be expressed by its dominating wavelength, colorimetric purity and luminosity. Two colour stimuli are compared by examining them side by side (simultaneous comparison), whereas two colour sensations must be compared successively, for it is only then that the influence of the surroundings upon the state of chromatic adaptation of the eye becomes fully manifest. Owing to chromatic adaptation the variations in colour sensation accompanying the changing over from one kind of light to another are much smaller than the differences in colour stimuli brought about by the two kinds of light. This phenomenon greatly affects the impressions we get from our surroundings.

Introduction

It is a common experience that a white tablecloth gives a "white" impression not only in daylight but also at night under the light of incandescent lamps. If this is considered more closely it is found to be very remarkable. In fact if an experiment is made where one part of the tablecloth is lighted by lamplight and another part by daylight — an experiment which anyone can improvise in the living room — the artificially lighted part will not appear to be white at all, but distinctly yellowish compared with that part under daylight. The spectral composition of the light striking the eye from the lamplighted part of the cloth (when it evokes a yellowish sensation) is, however, the same as that of the light striking the eye from the same cloth under the same lamp at night (when it is seen as white).

If the same experiment is made with coloured objects they also show unexpected differences in colour. A violet dress, for instance, will appear to be pinkish where it is under the lamplight, as compared with the part that is under daylight. In itself this observation is not very surprising, for the dress reflects light of different spectral distributions, corresponding to different colours according to the spectral composition of the two kinds of light falling upon it. The remarkable fact, however, is that if the room is lighted exclusively with electric light the colour of the dress is not pink at all, but violet.

These differences in colour only occur in everyday life when conditions are such that we have in the room daylight and electric light of about equal intensities but striking the objects from different sides. One then speaks of "false light". We then observe much greater differences in colour (*i.a.* coloured shadows) than the differences in shade which we are accustomed to observe when daylight is replaced by artificial light. That is why these differences in colour are often unexpected and disagreeable.

In order to explain the contrast between our observations in the experiments previously mentioned and our experience in everyday life, where no great differences in colour are noticed, we have to distinguish between the conceptions "colour stimulus" and "colour sensation". A colour stimulus is the subject of technical colour measurements. It is entirely different from the colour sensation, which is sometimes indicated in the following by the word colour, as applying to the subjective sensation of the observer. The fact that a given colour stimulus may give rise to different colour sensations is ascribed to a change in the retina under the influence of the light. This phenomenon is termed "chromatic adaptation" and it greatly affects the impressions we obtain from our surroundings, as is apparent from the foregoing. The object of the present article is to formulate clearly the difference between the conceptions "colour stimulus" and "colour sensation", preparatory to dealing with chromatic adaptation in a following article.

Difference between colour sensation and colour stimulus

In the following we shall explain the difference between the conceptions "colour sensation" and "colour stimulus". To that end the characteristic features of the two conceptions will be enumerated

in succession, each one first for the sensation and then for the stimulus.

Colour sensation: The colour sensation we get from an object in our surroundings depends on the following three groups of causes:

a) The spectral composition of the light that the coloured object throws upon the eye.
b) The "normal" laws of additive colour mixing of the eye, *i.e.* the laws governing the results of additive mixing of coloured light for the normal eye under standardised conditions.
c) All sorts of incidental circumstances affecting the state of our organs of sight at the moment.

Colour stimulus: The result of a technical measurement of colour is called colour stimulus, which only depends upon the factors a) and b). Methods of technical colorimetry have already been discussed in this periodical [1]).

Colour sensation: The circumstances c) affecting the colour of a beam of light of given physical properties come under the following headings:

1) Characteristics of the eye of the individual observer.
2) Properties of the objects viewed which evoke psychical influences (*e.g.* memory).
3) The state of the retina, which is affected by other light impinging upon other parts of the retina while the beam from the object is under view, or by such other light as may have just previously reached it.

Colour stimulus: In colorimetry the effect of these circumstances is precluded by various specific measures, such as:

1) The observer must have normal colour vision and the measurements must be performed with sufficiently high brightness [2]).
2) Essentially the measurement consists in the judging of the possibility to distinguish the two halves of a circular spot of light, one half having the colour to be measured and the other half a reference colour.
3) The rest of the visual field is dark.

Colour sensation: It might be considered ideal to possess a complete set of specifications for predicting the nature of a colour sensation from the various physical conditions. As, however, a colour sensation is difficult to express in numerical terms and, moreover, depends upon so many circumstances, some of which are of a non-physical nature, such an ideal can never be fully realised.

Colour stimulus: Thanks to the simple normal laws of additive colour mixing of the eye it has been possible to draw up complete specifications for the measuring of colour stimuli, so that nowadays there are tables enabling one to calculate a colour stimulus from the results of purely physical measurements, without any recourse to visual judgement. Colour stimuli are defined by means of the colour coordinates [3]) X, Y and Z (C.I.E. 1931).

Colour sensation: The characteristic features of a colour sensation are:

1) Hue: the property of colour sensation causing us to give the colour a name, such as red, green and blue.
2) Saturation: the extent to which a colour sensation differs from "white", or to which the sensation is "coloured"; the property that causes us to speak of faded colours or of vivid colours.
3) The impression of brightness: the property that causes us to speak of light and dark colours.

Colour stimulus: Instead of using the coordinates x, y and z a colour stimulus may be given by:

1) The dominating wavelength λ_d
2) The colorimetric purity p: a colour stimulus is reproduced by a mixture of the spectral colour λ_d and "white", the ratio of brightness being p: $(1-p)$.
3) The brightness, arising from the sensation of brightness by standardisation of the measuring conditions.

---

1) J. P. Bouma: The Perception of Colour, Ph. Tech. Rev. 1, 283, 1936.

2) Below a certain level of brightness the characteristics of the normal eye depend upon the brightness (Purkinje effect). See, *e.g.*, P. J. Bouma: The definitions of brightness and their importance in road lighting and photometry, Ph. Tech. Rev. 1, 142, 1936.

3) See *e.g.* P. J. Bouma: The representation of colour sensations in a colour space diagram of colour triangle, Ph. Tech. Rev. 2, 39, 1937.

This procedure is similar to the manner in which the conception of colour stimulus arises from the colour sensation.

Colour sensation: When judging a colour sensation one usually considers each colour separately, and if comparison is needed another colour is considered afterwards, so that the comparison is successive.

Colour stimulus: In visual colorimetry, which is the foundation upon which the aforementioned specifications are drawn up for determining the colour coordinates pertaining to a given colour stimulus, always two coloured patches of light are examined simultaneously: the stimulus to be measured and the reference colour stimulus, so that in this case a simultaneous comparison is made.

Colour sensations and colour stimuli due to different kinds of light

Bearing in mind the differences enumerated above, it is easy to understand the difference between the determination of colour sensations and the measurement of colour stimuli. The phenomena occurring when white or coloured objects are illuminated with different kinds of light also become clear. We will now proceed, in the same way as before, to compare in more detail the behaviour of colour sensations and colour stimuli due to different kinds of light.

Colour sensation: The question in how far the same colour impressions are obtained from a given coloured object under two kinds of light of different spectral composition is of importance when the whole of our surroundings is illuminated first with one of the sources of light and then with the other.

Colour stimulus: The question in how far the same colour stimulus is obtained from an object in our surroundings under two different kinds of light is particularly of importance when one desires to use both sources of light simultaneously.

Colour sensation: Changes occurring as a consequence of the transition from one kind of light to another may consist, *i.a.*, in a change in hue of some objects, or in a colour becoming more prominent owing to stronger sensations of brilliancy or of saturation. Such changes are quite marked when for instance incandescent light is replaced by mercury light.

Colour stimulus: Differences between two kinds of light present at the same time manifest themselves, *i.a.*, in the occurrence of the unpleasant phenomena of false light and coloured shadows referred to in the introduction. These phenomena are observed when an electric lamp is switched on in a room already illuminated by daylight.

Colour sensation: The equivalence of two kinds of light in respect to the colour sensations they evoke can as yet only be judged experimentally.

Colour stimulus: For the examination of the equivalence of two kinds of light as regards the colour stimuli they evoke for a given object there are, among others, two methods previously described in this periodical [4]); the first is more of an experimental nature and the second more theoretical.

Colour sensation: In essence the experiment consists in the examination of a large number of coloured cards against a white background first under one kind of light and then under the other. The comparison of the colour sensation obtained, as far as hue is concerned, is relatively simple because in this respect our impression can be adequately expressed in words ("I call this colour impression yellowish-green", etc). On the other hand it is much more difficult to compare the other two characteristics of the colour sensations (brightness and saturation). Therefore, the experiments to be described in a subsequent article will only have reference to hue denomination.

It is to be noted that here we have a successive comparison, where, after the change from one source of light to the other, the surroundings also throw a different light upon the eye and fully exercise their influence. This fits with the fact that we are dealing with colour sensations.

4) P. J. Bouma: Colour reproduction in the use of different sources of "white" light, Ph. Techn. Rev. 2, 1, 1937.

Colour stimulus: In the experimental method a number of coloured cards are held against a black background and one half of each is exposed to one of the kinds of light while at the same time the other half is exposed to the other light. The difference in colour stimulus (also with regard to luminosity and colorimetric purity) showing itself in the simultaneous examination of the two halves is a measure for the difference between the two sources of light.

Here we have a simultaneous comparison and the conditions of measurement are approximately those of colorimetry. This fits with the fact that we are dealing with differences in colour stimuli.

Colour sensation: A more theoretical method of comparison based on the spectral composition curves of the sources of light can only be developed if the influence of the surroundings appears to be subject to definite laws. Then the influence can be accounted for and predictions can be given as to the changes that will take place in the colour sensations evoked by a coloured object when changing over from one kind of light to another. This, too, will be gone into more fully in a subsequent article.

Colour stimulus: The second, more theoretical, method of comparison already mentioned [5]) is based on the spectral composition curves of the sources of light. The spectrum is judiciously divided into 8 sections. It may then be said that the two sources of light are to be regarded as being practically equivalent if the relative contributions of each of them to the luminous flux in every section is the same within certain tolerances. Where there is a noticeable difference between the sources of light the degree of that difference can be judged by calculating the differences in trichromatic specifications that will occur for the coloured object when changing over from one of the kinds of light to the other.

The examples quoted in the introduction have already given some idea how great the differences may be under two different kinds of light, and this can be further illustrated in the following way. The colour points of the reflected light for 24 cards of a "circle" of an Ostwald colour atlas are calculated when exposed successively to daylight and to the light from an incandescent lamp [6]). We selected for this purpose a circle of rather saturated colours, the *nc* circle, consisting of 100 cards in all, of which the saturation impression is fairly con-

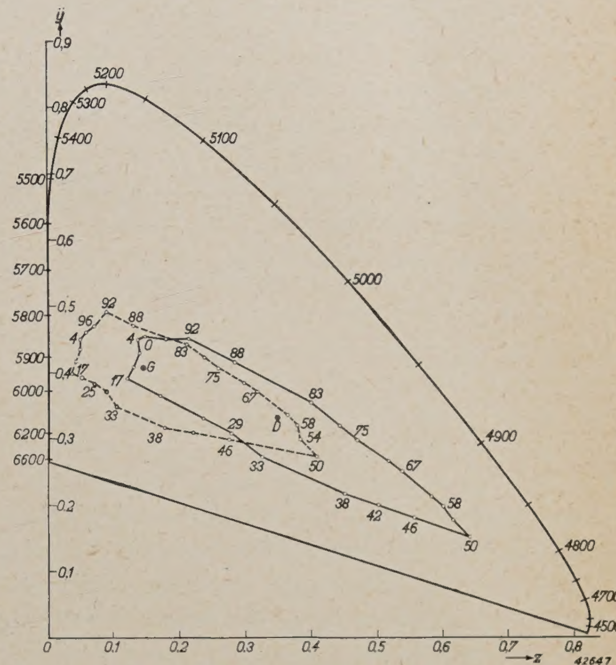


Fig. 1. Position of the colour points for the Ostwald cards *nc* O, 4, 8, 13, ... in the international colour triangle. Broken line: under the light from an electric incandescent lamp. Full line: under daylight. *D* is the colour point for daylight, *C* for incandescent lamplight, *y* and *z* are colour coordinates of the C.I.E. The numbers from 4500 to 6600 indicate the wavelength in Ångstrom units on the curve of the spectral colours.

stant. The hue of the cards numbered 0 to 99 extends from yellowish-green to yellow, orange, red, purple, blue, green and back to yellowish green [7]). The results of the calculations made are given in *fig. 1.*

Colour sensation: When, for instance, card 83 is first examined under daylight and then under electric light, the colour appears to be practically the same: under both kinds of light this card gives a purely green sensation.

Colour stimulus: Fig. 1 shows, however, that the two colour points 83 are far apart: the colour stimuli are different. This difference in colour stimulus is obvious when one half of the card is exposed to daylight and the other

[5]) See also P. M. van Alphen: A photometer for the investigation of the colour rendering reproduction of various light sources, Ph. Techn. Rev. 4, 66, 1939.

[6]) For the method of calculations see Philips Techn. Rev. 2, 45, 1937.

[7]) In the new editions of the Ostwald atlas the circle has only the 24 cards used here.

half to electric light and both are examined simultaneously.

Colour sensation: If one examines in succession card 47 under electric light and card 31 under daylight a large difference is noticed, the former showing a violet hue and the latter crimson.

Colour stimulus: Fig. 1 shows that the two colour stimuli, 47 under electric light and 31 under daylight, are practically the same. This is confirmed by a test similar to that with the two halves, holding side by side card 47 under electric light and card 31 under daylight (simultaneous comparison).

The results of these experiments are to be explained by the fact that when comparing colour sensations we find the results are influenced by the activity of an additional factor, *viz.* a factor classified under c) in the beginning of this article. The experiments show that this factor may contribute towards making two equal colour stimuli appear to be different in the sensations they evoke (the last example), but that on the other hand, under certain circumstances, it may neutralise a difference in colour stimuli in so far as the colour sensation is concerned (first example). In these cases this factor is mainly to be sought in the surroundings.

Colour sensation: The surroundings usually have a considerable influence upon colours, as is evident from the last mentioned example where in different surroundings one and the same colour stimulus gives an entirely different sensation. This influence lies mainly in the fact that the retina gets a different sensitivity for the various colours. This phenomenon is called chromatic adaptation of the eye.

Colour stimulus: The surroundings do not affect the colour stimuli. Though the specifications for colour measuring prescribe entirely dark surroundings, any non-dark surroundings have practically no influence upon the result, because the two coloured patches to be compared are always shown simultaneously and immediately adjacent to each other; consequently the two parts of the retina on which the images of the coloured patches are formed are always adapted in approximately the same way.

The great influence of the surroundings upon colours can be further demonstrated in the following way (see *fig. 2*). A transparent window 15 × 15 cm can be illuminated at the back with a number of differently coloured lamps. Around the window is a field of 80 × 80 cm which may emit incandescent lamplight or artificial daylight as desired, without affecting the light passing through the window.

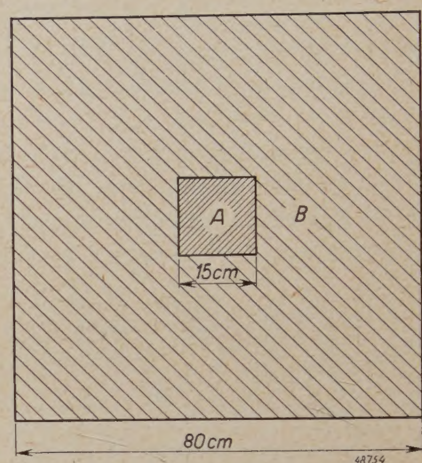


Fig. 2. The chromatic adaptation of the eye is very well demonstrated by illuminating a transparent window *A* with a light of a certain colour and the surrounding field *B* first with artificial daylight, for instance, and then with incandescent lamplight. Shortly after changing over from daylight to lamplight the eye directed upon *A* sees a change of colour in the window.

If the whole set-up is viewed from a few yards away and the light in the surrounding field is changed from daylight to incandescent lamplight then in most cases after a while a decided change of colour seems to take place in the small lighted window.

Colour sensation: The colour sensation may change from say orange-yellow to yellowish-green, or from crimson to bluish-purple. In a few cases, however, the colour does not change at all (in the cases of blue and orange).

Colour stimulus: The colour stimulus that can be measured at the window does not, of course, change because the lamps are still the same. Thus we see here a change in colour sensation due to the influence of the surroundings in its unqualified form [8]).

In the example of card 83 we saw that it may happen that under certain circumstances a change in colour stimulus does not bring about a change in colour sensation: the process of chromatic adaptation

---

[8]) The fact that the change in colour is not noticed until some time after the surrounding light is changed indicates that the eye requires some time to adapt itself to the changed surroundings.

of the eye compensates a difference in stimulus arising in the beginning. This phenomenon occurs in practice (consider the examples in the introduction) so frequently that it has been formulated in a "law", which for a number of years already has been widely adopted by psychologists and physiologists, *viz*:

Colour sensation: "The colour sensations created by the coloured objects in our surroundings are practically independent of the kind of light with which the whole scene is illuminated".

Colour stimulus: At the same time, however, the colour stimuli may differ appreciably, as illustrated in fig. 1.

The aforementioned rule of the unchangeability of colour sensations, however, holds only for kinds of light having a spectral composition differing little from that of black body radiators, and even so it is only an approximative rule. Small deviations can easily be observed experimentally. For instance, pigments having a different spectral remission curve may have the same colour when exposed to one certain kind of light but they are certainly different after the light has been changed.

We therefore prefer to formulate the rule as follows: *With most kinds of light commonly used the difference between the colour stimuli going out from a given pigment under two different kinds of light is usually much greater than the difference between the colour sensations evoked by those colour stimuli.*

For example, when comparing the light from an incandescent lamp with daylight:

Colour sensation: Everyone knows from experience that there is very little difference in the colours of the same object when illuminated by these two sources of light.

Colour stimulus: From fig. 1, or when applying the "8-sections method" or "the comparative method of two halves" (see the foregoing), the conclusion to be drawn is that the difference in colour stimuli for a given card under different light may in fact be surprisingly great.

Here we may conclude the comparison of the two conceptions, colour stimulus and colour sensation. It has already been pointed out that much less is known about colour sensations than about colour stimuli. In particular very few investigations have as yet been made in regard to the determination of colour sensations produced under different circumstances. It is hoped that the experiments and observations to be dealt with in the next article will contribute towards our knowledge of this subject.

A REMARKABLE PHENOMENON WITH STEREOPHONIC SOUND REPRODUCTION

by K. de BOER.

To a practised listener the sound image heard with stereophonic reproduction generally appears to lie above (and at times below) a line between the loudspeakers, in other words it has elevation. This remarkable phenomenon can be explained in every detail, even quantitatively, by the familiar hypothesis that the impression of elevation of sound above the horizontal plane of its source is due to slight movements of the head. For stereophonic reproduction this phenomenon is of no consequence in practice.

When a system for stereophonic sound reproduction is installed, such as already described in this journal, [1]) two loudspeakers are set up on either side of a platform or screen, each with its own channel (amplifier, etc.) from a separate microphone set up in the recording room (*fig. 1*). For cinema reproduction the sound striking each of the two microphones in the film studio is first recorded on a sound track on the film and each of the loudspeakers reproduces the sound from the corresponding sound track. This case, too, can be represented diagrammatically by fig. 1.

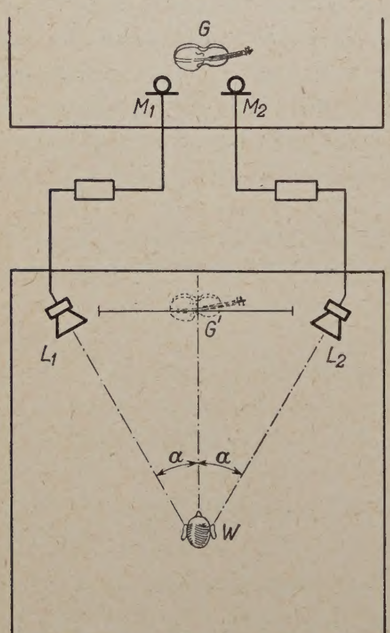


Fig. 1. Schematic representation of a set-up for stereophonic reproduction. Two loudspeakers L_1 and L_2 set up in the reproducing space B are each connected to a corresponding microphone M_1 and M_2 in the recording space A. Here, for the sake of clearness, the microphones are shown separated, but in reality they are usually contained in a dummy head. The listener W observes a virtual picture G^1 of the source of sound G. The recording space may be a film studio and the reproduction space the auditorium of a cinema, in which case the "channel" between each microphone and its corresponding loudspeaker comprises, *inter alia*, a sound track in which the proportions of sound are preserved.

Let us suppose that a listener is standing in the middle of a hall where the loudspeakers are set up. Particularly when the spoken word is being reproduced, a "sound image" can very well be localised, that is to say the sound is heard to come from a certain spot between the two loudspeakers, depending on the position of the source of the sound in the recording room. For the present we will confine our attention to the case where the sound image appears to lie just halfway between the loudspeakers.

With the help of a number of trained listeners the astonishing fact has been discovered that as a rule the sound image is not situated on the line running direct from one loudspeaker to the other, as one would expect, but a distance above it [2]): it has a certain elevation above the horizontal plane. Upon walking down the middle of the hall towards the loudspeakers the listener hears the sound coming from a higher level, rising gradually at first but more quickly as he gets nearer, until when he has reached the centre of the line between the loudspeakers the sound is practically perpendicular above him.

Still more astonishing is the sensation when the listener raises his head to try to "see" whence the sound is coming, as one does when there is some source of sound with elevation (*e.g.* an aeroplane). The sound image seems to climb higher and higher, so that one cannot catch it in the eye, as it were.

Some observers find that when they concentrate their minds and keep looking straight ahead the sound does not come from above but rather from below the horizontal plane, while as they move closer forward it goes deeper and deeper, finally disappearing through the floor, vertically underneath. To some this downward effect comes more readily than the upward movement.

In practice, with stereophonic reproduction in a concert hall or cinema such an effect has no

[1]) K. de Boer, Stereophonic Sound Reproduction, Philips Techn. Rev. 5, 107, 1940.

[2]) K. de Boer, A remarkable phenomenon in direction of hearing, Ned. T. Natuurk. 11, 75, 1944.

adverse consequences worth mentioning. From most seats in the auditorium the distance from the loudspeakers is so great that the elevation of the sound can only be small; something further will be said about this in the following pages. Moreover, on the one hand most of the seats are not in the middle of the hall, while on the other hand the sound image lies (or the images lie) closer either to the right-hand loudspeaker or to the left one, both tending to reduce the elevation observed. Furthermore, this is by no means an effect that strikes everyone. In fact it takes some practice to notice it at all, as is understandable considering that in general it is much more difficult to observe and determine the elevation of a sound than the azimuthal angle. As a consequence the observation of the elevation of a sound image is more susceptible to suggestion. In the case, for example, of stereophonic reproduction of the sound recorded of a passing aeroplane, all listeners in different places seemed to hear the sound directly overhead. It is just by reason of this susceptibility to suggestion that where stereophonic reproduction of sound accompanies picture projection (in a cinema) the sound image is always attracted by the visual picture towards the horizontal plane, even if a listener is sufficiently trained to observe an elevation effect. A similar suggestive influence is present in the case of reproduction of a musical performance, for then the listeners picture to themselves the places usually occupied by the various members of the orchestra.

Although this effect has no practical consequences, it is instructive to discuss it and to show how the phenomenon can be explained both qualitatively and quantitatively. It is to be remarked that it can be observed at home, for instance when in order to improve the quality of one's radio set ("smoothing out" the sound) one has two loudspeakers set up some distance apart [3]).

Description of the experiments and results of measurements

In order to investigate the effect described and in particular to measure the apparent elevation observed, an arrangement was set up as shown in *fig. 2*, with two loudspeakers at the level of the listener's ears and halfway between them a vertical measuring rod. The listener is placed in the plane of symmetry with the set-up and told to look straight ahead of him, at a mark on the measuring rod. When there are small angles of elevation of the sound image he can read their position directly

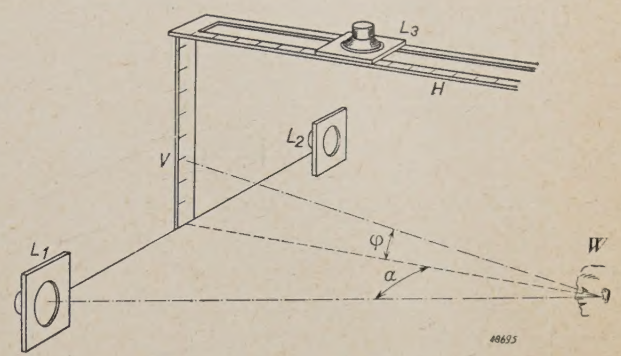


Fig. 2. Set-up for measuring the elevation effect observed. V is a vertical measuring rod on which the angle φ for small elevations can be read directly by the listener W. H is a horizontal rod with auxiliary loudspeaker L_3 that can be moved along it for determining larger elevations.

on the rod, without lifting his head (this he must not do, because then the image moves higher up). Because large angles of elevation cannot be read directly from the vertical measuring rod, another was set up horizontally, along which an accessory loudspeaker could be moved (fig. 2). When this loudspeaker is in a suitable position the sound from the apparent source to be localised seems to come from the same direction as a momentary signal emitted at intervals from this loudspeaker. In this way the direction of the sound image is determined by a "zero method"; the observer has only to judge whether there is any difference in direction in respect to a comparative signal.

The results of the measurements are reproduced in *fig. 3*. They were obtained for the greater part with little trained listeners. With the difficulty already mentioned in this kind of localisation the observations were consequently not very consistent and showed considerable variations. Nevertheless, the diagram shows clearly the qualitative details of the phenomenon. The angle of elevation φ of the sound image observed is plotted for different distances from the listener to the pair of loudspeakers, measured through the angle α between the plane of symmetry and the line from the loudspeaker to the listener (figs. 1 and 2). It is seen that as the angle α becomes larger and thus the listener gets nearer the sound image rises first slowly and then more quickly. At $\alpha = 90°$, that is to say when the observer is just between the loudspeakers, the

[3]) When two loudspeakers are used, either connected to the same radio set or placed for stereophonic reproduction, in order to get a natural impression with a well defined sound image it is necessary that the two loudspeakers should be in phase with each other, that is to say the diaphragms must both move together towards and away from the listener. One loudspeaker is put in phase or counter-phase with the other simply by turning it or changing the polarity of its connection.

sound does not come from exactly overhead but from a short distance behind ($\varphi \approx 100°$).

Explanation of the phenomenon

Since the phenomenon described consists of an elevation of an apparent source of sound, obviously

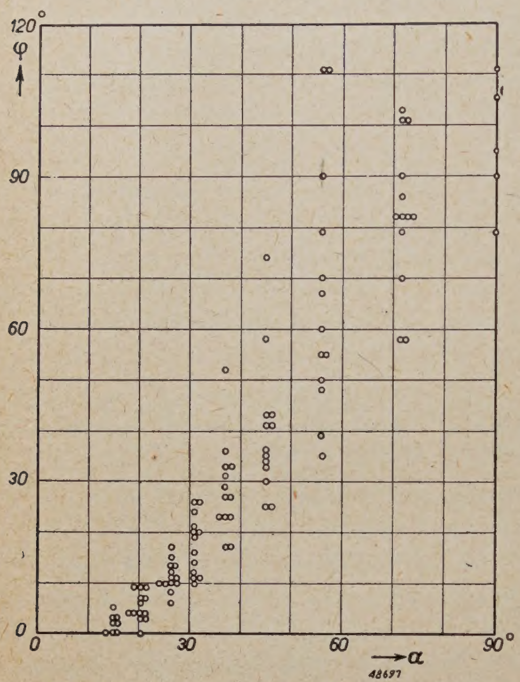


Fig. 3. Measured values of the angle of elevation φ observed by a number of (little trained) listeners, as a function of the distance between the listener and the pair of loudspeakers (measured through the angle α, see figs. 1 and 2).

its explanation must be related to the theory accounting for the perception of the elevation of an actual source of sound.

This theory has already been dealt with *in extenso* in this journal [4]). Briefly it comes to this, that a perception of direction is due to a difference in intensity between the two ears [5]). Where there is no such difference, *i.e.* where the intensity ratio $v = 1$, this corresponds to the perception of direction "straight ahead". Equally so, however, $v = 1$ happens with a source of sound "immediately overhead" or "immediately behind", or in general with any elevation of the source of sound in the plane of symmetry with the observer (*fig. 4*). A criterion for distinguishing these cases and "determining" the angle of elevation is obtained by the listener making small more or less arbitrary turns of his head around a verticale axis (as shaking the head for "no"). When giving the head a turn dα the ratio v of the intensity between the ears will change by an amount dv approximately proportional to dα. In particular, for instance when the source of sound is "straight ahead", when the plane through that source of sound and the aural axis is horizontal, dv = 3% (0.14 db) for dα = 1°. If, however, the source of sound has an elevation φ, then for a given turn dα the effective turn dα' in the plane through the source of sound and the aural axis (fig. 4) is smaller, *viz.*

$$\mathrm{d}\alpha' = \mathrm{d}\alpha \cdot \cos\varphi .$$

Also the change dv' in the intensity ratio is proportionately smaller for a turn dα. Now the fact that for a given dα a too small dv' occurs is interpreted by the sense of hearing (due to its experience) as an elevation of the source of the sound, and it does this quantitatively, according to the equation[6]):

$$\cos\varphi = \frac{\mathrm{d}\alpha'}{\mathrm{d}\alpha} = \frac{\mathrm{d}v'}{\mathrm{d}v} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The elevation thus "measured" by the ear is in reality not the angle in respect to the horizontal plane but one in respect to a fixed plane of reference connected with the head, which with the head in the normal position is horizontal. This is of importance in the event that the head is lifted, for the

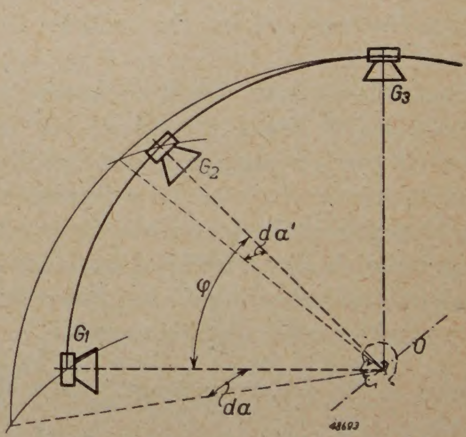


Fig. 4. Various sources of sound G_1, G_2 .... in the plane of symmetry with the listener's head all give an intensity ratio $v = 1$ between the sound at the two ears. When the head is turned (as when shaking for "no") through an angle dα a change dv takes place in the intensity ratio, which is greatest for the source G_1 in the horizontal plane. For a source G_2 with elevation φ the effective turn dα' in the plane through G_2 and the aural axis O is only dα cos φ.

---

4) K. de Boer and A. Th. van Urk: Some facts about the direction of hearing, Philips Techn. Rev. 6, 363, 1941.

5) Accompanying differences in time also contribute to a perception of direction (*cf.* the article quoted in footnote 1)), but for very small angles α they are of such little influence that we may here confine our considerations to the differences in intensity.

6) A similar, somewhat generalised equation applies to the case where the source of sound does not lie in the plane of symmetry of the observer. when the intensity ratio v differs from unity already in the state of rest. For the sake of simplicity, however, we confine ourselves to sources of sound in the plane of symmetry.

plane of reference moves with it. When the listener gets the source of sound in line with the eye it is then again in the plane of reference and he therefore expects his sense of hearing to register the elevation zero.

So much for the explanation of the perception of the elevation of an actual source of sound. We will now go back to our set-up for stereophonic reproduction, fig. 1. There we have an apparent source of sound observed straight ahead as the result of the joint action of two loudspeakers, one on the right and one on the left. We can again put the question what change will take place in the intensity ratio at the ears when the listener turns his head an angle da from the vertical axis. *It is found that this change is in fact smaller than what would be expected with an actual source of sound straight ahead.*

The fact that the dv' will be smaller than for an actual source of sound may be seen from the following:

We will use I_1 and I_2 to denote the sound intensities from one loudspeaker at the ear closest to it and at the other ear respectively, when the loudspeaker is placed at an angle a from the plane of symmetry (*fig. 5a*). When the listener looks straight ahead each ear receives the total intensity $I_1 + I_2$. Upon the head being turned an angle da to the right the angle of deviation of the loudspeaker L is increased to $a + da$ while that of the loudspeaker R is reduced to $a - da$. The left ear then receives (fig. 5*b*) the quantities of sound

$$I_1 + \frac{dI_1}{da}\,da \text{ from } L, \text{ and } I_2 - \frac{dI_2}{da}\,da \text{ from } R.$$

The same applies for the right ear. The intensity ratio at both ears, which was originally unity, is now:

$$1 + dv' = \frac{I_1 + \dfrac{dI_1}{da}\,da + I_2 - \dfrac{dI_2}{da}\,da}{I_1 - \dfrac{dI_1}{da}\,da + I_2 + \dfrac{dI_2}{da}\,da}$$

Thus $1 + dv'$

$$\approx 1 + \frac{2}{I_1 + I_2} \cdot \frac{d}{da}(I_1 - I_2)\,da.$$

$$\text{or } dv' = \frac{2}{I_1 + I_2} \cdot \frac{d(I_1 - I_2)}{da}\,da \quad . \; . \; . \quad (2)$$

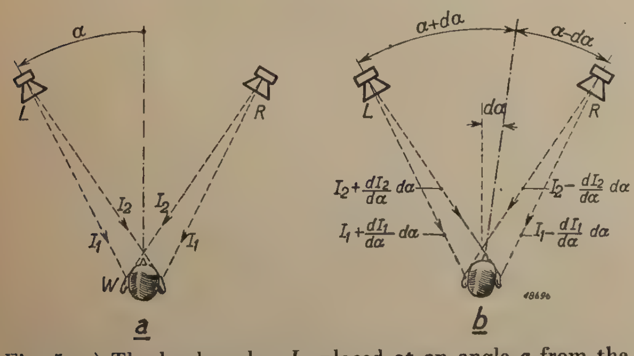


Fig. 5. *a*) The loudspeaker L, placed at an angle a from the plane of symmetry of the head of the listener W, produces the sound proportions I_1 and I_2 respectively at the ear closest to it and at the other ear. The same is the case with the loudspeaker R (the sound image is presumed to lie in the plane of symmetry). *b*) After a small turn of the head da the left-hand loudspeaker is at an angle $a + da$ from the plane of symmetry of the listener, while the right-hand one is at an angle $a - da$. The respective sound proportions at each ear have been altered accordingly, to the values indicated in the diagram.

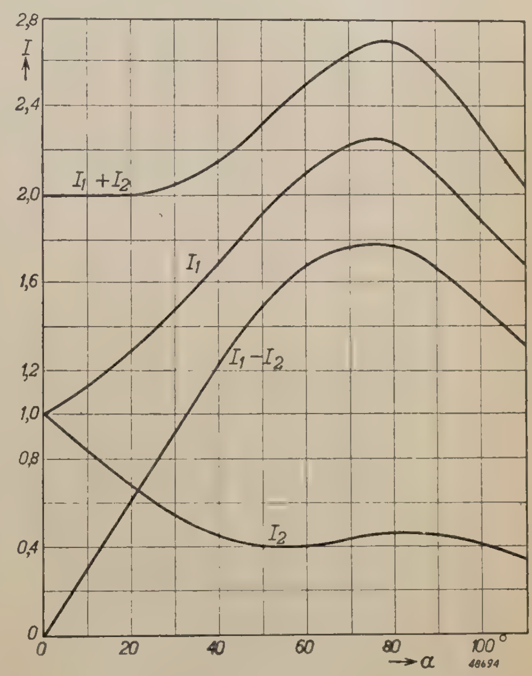


Fig. 6. The sound proportions from a loudspeaker at the closest ear (I_1) and at the other (I_2) as functions of the angle a through which the loudspeaker is turned away from the plane of symmetry of the listener (see the arrow at a in fig. 5*a*). The curves have been plotted from measurements taken by Sivian and White [7]). There are also plotted the quantities $I_1 + I_2$ and $I_1 - I_2$ used for calculating the elevation observed.

The functions $I_1(a)$ and $I_2(a)$ indicate simply the sound intensities from one single loudspeaker, say L in fig. 5*a*, at the ear closest to it and at the other ear respectively when that loudspeaker is moved in respect to the listener's head in the manner indicated by the arrow in fig. 5*a*. This intensity as a function of a has been dertermined by Sivian and White [7]), whose results are reproduced in *fig. 6* by the curves I_1 and I_2. This figure also gives the curves $I_1 + I_2$ and $I_1 - I_2$ as a function of a. By this means it is possible to calculate the effective change dv' in the intensity ratio according to (2) for any

---

[7]) J. Sivian and S. B. White, J. Acoust. Soc. Amer. 4, 288, 1933.

angle a (corresponding to a certain distance between the listener and the pair of loudspeakers) and then to predict the "elevation" to be expected according to (1) by comparison with the "normal" change already given ($dv \approx 3\%$ for $da = 1$).

The fact that $dv' < dv$, thus that there will actually be a real angle of elevation ($\cos \varphi < 1$), may be deduced from fig. 6, considering that in the limit case $a = 0$, when the listener is removed far from the loudspeakers, the two loudspeakers together function as one actual source of sound straight ahead. The change dv' calculated from (2) for this case will therefore be the same as that to which the ear is accustomed and used as criterion (dv) in normal hearing in such a situation as this. From fig. 6 it will now be seen that the curve I_1—I_2 (denominator of the fraction in (2)) lies lower for $a = 0$ than for any other value of a, whilst the slope of the curve I_1—I_2 (numerator of (2)) is greater for $a = 0$ than for any other value of a. The change observed (dv') is thus certainly smaller for any angle a than that for $a = 0$, *i.e.* the angle normally expected.

Further proof of the explanation

The calculation as described of the elevation to be expected with the aid of the curves in fig. 6 and the equations (2) and (1) gives as a result the curve plotted in *fig.* 7, which shows also the elevations determined experimentally, *viz.* the mean values of the series of measurements reproduced in fig. 3. It will be seen that there is a fairly good agreement. In the range of the small angles the calculation is very inaccurate, due to the not inconsiderable influence of the limited accuracy of Sivian and White's measurements (which, moreover, depend somewhat upon the frequency spectrum of the sound). For large angles a the calculation is more reliable and even leads in fact to a prediction of the peculiarity, already mentioned, that as the listener reaches the line connecting the two loudspeakers the sound comes from immediately behind him instead of directly overhead. It is certainly remarkable that the simple theory developed here should produce also this detail.

The fact that some listeners may observe a negative elevation instead of a positive one is not contradictory to the theory, because since a too small dv' occurs just as well with a negative elevation as with a positive one the theory makes no differentiation between these two cases. Why, however, in some cases there happens to be a preference for the perception of the rather unusual negative elevation is difficult to account for.

There remains to be considered the impossibility of getting the apparent source of sound in the line of sight by lifting the head. This, too, is not difficult to comprehend. If, when listening to the real source of sound with an elevation, one gradually raises the head, the change dv' in the intensity ratio at the ears becomes greater and greater when the head is moved to either side, owing to the elevation of the source above the plane of reference decreasing.

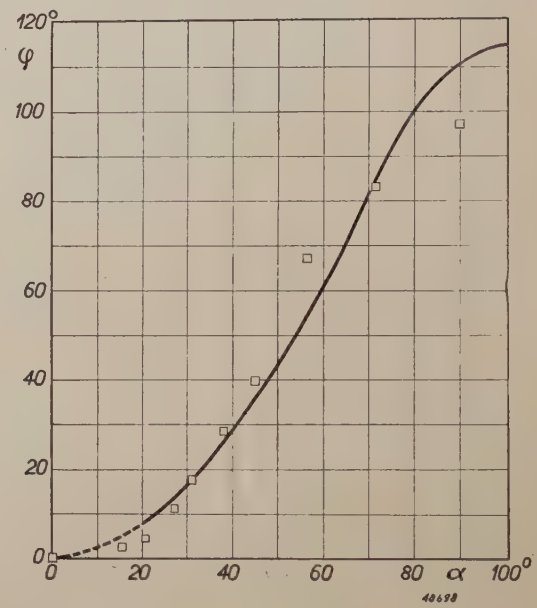


Fig. 7. Calculated elevation φ of an apparent source of sound straight ahead, as function of the distance to the two loudspeakers (angle a). For very great distances (small angle a) the calculation is rather unreliable; the curve for this range is drawn in a broken line. The squares are the mean values of the series of measurements given in fig. 3.

When the plane of a reference passes through the source, dv' becomes equal to the value that the sense of hearing regards as normal, thus no longer "too small", and the impression of an elevation disappears. In our experiments with an apparent source of sound the "too small" value of dv' was caused by the joint action of the two loudspeakers. Upon the head being raised, while still looking straight ahead, the proportions of sound at the two ears, including their changes when shaking the head, remain approximately the same, so that dv' continues to be "too small" and the impression of elevation remains. Consequently the sound image rises at the same time.

The more complicated phenomena occurring when the apparent sound image does not lie halfway between the loudspeakers or the listener is not in the plane of symmetry can also be explained in the manner described. We shall not enter into these cases here. We would only recall the fact already

mentioned that in such cases the elevation is always smaller than in the case we have considered above. By this means and with the aid of fig. 7 we can give an indication of what significance this effect might have in stereophonic reproduction in practice, for instance in a cinema. Any perceptible elevation of about 20° will easily be corrected by the suggestion of the visual picture, so that no elevation will be noticed in the seats farther away from the screen than the place corresponding to the angle $a = 30°$ (see fig. 7). If the distance between the loudspeakers is say 8 meters the elevation effect will therefore only be noticeable for a trained listener at places closer than about 7 meters to the screen, and the number of such seats — which also in other respects are to be regarded as unfavourable — is only a small percentage of the total number of seats in a normal cinema.

A NEW ELECTRICAL METHOD FOR DETERMINING MOISTURE CONTENT

by J. BOEKE. 543.812: 621.317.39

A description is given of a new method for determining the water content of liquid or solid substances, such as grain, wood, textiles, butter, etc. The water is extracted from the material with acetone in which oxalic acid is dissolved. The increase in the conductivity of the extraction liquid serves as a measure of the amount of water taken up. Compared with those already existing this method has the advantage that the apparatus required is inexpensive and simple in operation, while at the same time it is a very reasonably quick method and not dependent on the form in which the substance to be tested occurs.

The determination of the water content in gases, liquids and solids is one of the most important analyses regularly applied in commerce and industry. The moisture content of grain, of tobacco, of butter, to name only a few examples, is also paid for by the purchaser, and large profits or losses may be involved when a kilogram of the product contains a few grams more or less of water than were estimated. It is therefore understandable that numerous attempts have been made to develop methods of measuring moisture content quickly and accurately.

If we confine ourselves to solids and liquids, the only universal method used until now consisted in drying a sample and determining the amount of evaporated water by weighing. This, however, is often a lengthy and far from easy procedure. More rapid work is made possible by a number of electrical methods where measurements are taken directly on a sample without any preliminary treatment. Three main methods of measurement have been developed on this principle. In the first the difference in dielectric constant is determined between the moist and the dry substance; in the second the dielectric losses are measured; in the third the conductivity. In all these methods, however, the results depend very much upon the form of the substance to be examined (lumps, grains, powder, liquid), for this form affects the filling factor when the sample being measured is introduced into a measuring vessel. The result is that only relative values of moisture are measured, and if absolute quantities of water are desired a separate calibration curve is needed for each substance in its given form.

In order to eliminate the effect of the form of the substance (which means that the absolute water content is determined with only one calibration curve for all substances), another group of methods have been developed which are based on the extraction of the water from the moist substance. The most important methods of this type are tabulated and briefly characterized below. Except for the drying already mentioned, which actually also belongs to this group and is therefore included in the table, the methods indicated here are hardly less rapid than the direct electrical measurements, but they all have one drawback in that they require a fairly expensive apparatus which cannot be operated by a layman [1]).

We have now found that a method of measurement on the basis of extraction can be developed which while still being reasonably quick offers the advantages of very simple operation and an inexpensive apparatus.

The basic idea is the following. Water is extracted from the substance to be tested with a hygroscopic

[1]) A survey of the different methods of moisture determination will be found in: E. Eckert and P. Wulff, Z.angew. Chemie, 53, 403-405, 1940.

Process	Extraction medium	Procedure	Determination of the water extracted by:
Drying	air (vacuum)	circulation, or heating	weight
Distillation	toluene	distillation	volume
of mixture "Exluan" process	dioxane	mixing, or grinding together	dielectric constant
Titration according to Fischer	methanol	mixing, or grinding together	potentiometric titration with Fischer's reagent

liquid which in itself has a low conductivity and does not readily dissociate electrolytes. An electrolyte is dissolved in the extraction liquid, thereby being only very slightly dissociated in it, so that the solution possesses only a low conductivity. The absorption of water by the extraction liquid increases its power to dissociate, the electrolyte is thus more dissociated and the result is a considerable increase in the conductivity of the liquid, which can be measured by simple means.

As extraction liquids possessing the desired properties methyl and ethyl alcohol can be used, but acetone is still more suitable. Oxalic acid may be used as electrolyte, since it dissolves very readily in acetone [2]). When a solution of 10% oxalic acid in acetone is used the conductivity of the solution as a function of the water content increases very rapidly, as may be seen from the curve in *fig. 1*. Variation in the content of oxalic acid makes relatively little difference, as shown by the dotted-line curves; the effect of changes in temperature is also slight.

The possibility that the test substance contains common salt or some other salts must be kept in mind. Common salt is practically insoluble in pure acetone and has scarcely any effect on the low conductivity of acetone. A slight addition of water, however, causes the salt to dissolve more readily in the acetone, the dissolved salt is dissociated and the conductivity increased. This effect, which is based upon the increased dissolving power of the acetone upon absorption of water, is even stronger than the effect of the increased dissociating power, of which use is made in our method of measuring, and it may therefore obviously be asked why the whole method should not be based rather upon the first effect. That would mean that an excess of solid sodium chloride would have to be added instead of oxalic acid, but then there is the objection that sodium chloride dissolves very slowly in acetone containing little water (say < 5%), and it may take days to establish the equilibrium, which would be very objectionable for practical measurements. For our method, based on the dissociation of oxalic acid, on the other hand, the slow solution of sodium chloride has just the advantage that little difficulty is experienced from the salt content of the substance to be examined. It is only necessary that the extraction should be completed within two hours and that the final acetone-oxalic acid-water mixture does not contain more than a few percent

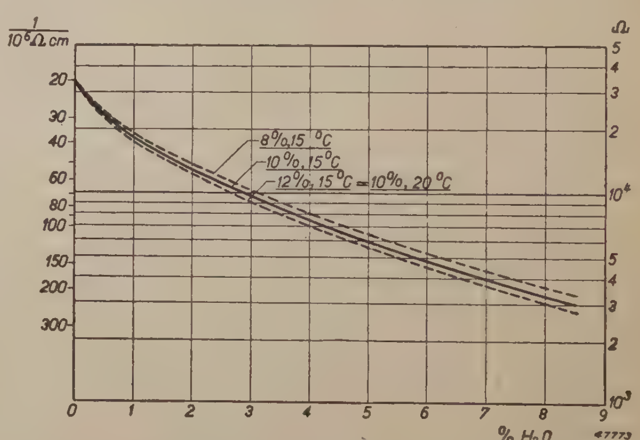


Fig. 1. Conductivity and resistance, measured in the measuring cell, of acetone with a certain percentage of dissolved oxalic acid, as a function of the amount of water absorbed. The three curves apply for different concentrations of oxalic acid and temperatures as indicated.

of water. Provided these conditions are fulfilled, large deviations from the calibration curve of fig. 1 are only observed in the case of substances with an extremely high content, such as salted fish. The apparent amount of water in such a case may be double the real amount determined by drying. With wood, paper, textiles and suchlike, however, the figures found for the water content, taken absolutely, differ by only a few tenths percent at most from those found by drying. The accuracy of the method, like that of the other methods based upon extraction, is for a large part limited by the familiar phenomenon that the test object very stubbornly retains the last traces of water and the state of equilibrium, where only a very minute quantity of water remains definitively in the sub-

[2]) It would be simpler if the extraction liquid itself could function as electrolyte; concentrated sulphuric acid for instance is hygroscopic and at the same time upon taking up water its dissociation and consequently its conductivity is very much increased. It has not been possible, however, to find a substance which combines all the desired properties. Sulphuric acid, for example, is too aggressive chemically.

stance, is only slowly attained. Substances containing albumins are especially difficult in this respect. Grains of wheat, for example, which had lain in a vacuum of 20 mm Hg for three days, while for 10 hours the temperature had been maintained at 65 C°, again lost 1.6% by weight of water during the next two days in a vacuum. This phenomenon makes it very difficult, whatever the method, to determine the absolute moisture content of such substances. Nevertheless, the extraction with acetone has at least the advantage that the above-mentioned equilibrium is established much more quickly than when drying in air or a vacuum.

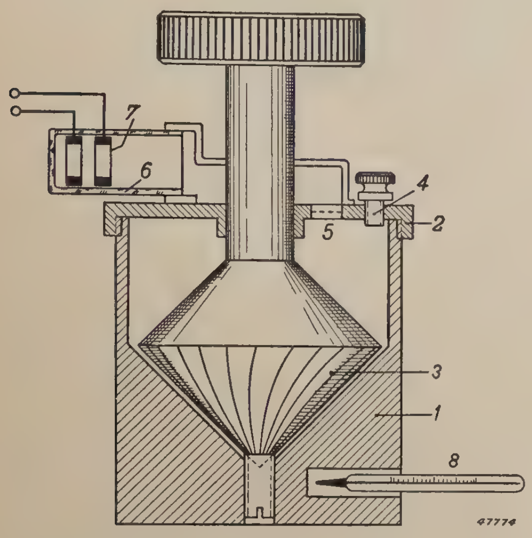


Fig. 2. Apparatus for determining the moisture content by the extraction of water and the measurement of conductivity. After the vessel *1* is filled with a weighed sample of the substance to be investigated, the cover *2* with the grinding cone *3*, which can be rotated and moved up and down, is screwed on. Through the opening *4* a known amount of extraction liquid is poured in. By rotating and pressing at the same time on the cone, which is provided with grinding grooves, the sample and the liquid are mixed. After the extraction is complete, the apparatus is tipped and the liquid runs through the sieve *5* into the measuring cell *6* which is attached with an air-tight joint and contains the platinum electrodes *7*. The resistance between these electrodes can be measured. *8* is a thermometer.

In the practical application of the method care must be taken, *i.a.*, that during the extraction and the subsequent measurement of the conductivity no acetone can evaporate, which would increase the concentration of the oxalic acid. It is therefore desirable to carry out the measurement in a completely closed vessel, as shown in *fig. 2*. The apparatus consists of a kind of mill with a measuring vessel attached to it. A weighed quantity of material and a measured volume of acetone with 10% oxalic acid are placed in the mill, which is then closed tight. After the substance and the extraction liquid have been thoroughly mixed together and left to stand for one to two hours, the mill is tipped so that the extract, filtered through a sieve, runs into the measuring vessel. This is a cell with two platinum electrodes. Since the configuration of the electrodes and the liquid between them is fixed, from a measurement of the electrical resistance of the cell the conductivity of the liquid and, with the help of the calibration curve in fig. 1, its water content can immediately be derived. The resistance may be measured by means of a measuring bridge, for example the "Philoscope" [3]), which is specially adapted for such simple measurements. In the choice of the calibration curve the temperature of the extract, which can be read off on a thermometer attached to the mill, must be taken into account.

Since oxalic acid attacks metals, all the surfaces of the mill coming into contact with the extraction liquid, except the platinum electrodes, must be made for instance of ceramic material, plastics or glass.

[3]) See Philips Techn. Rev. 2, 270, 1937.

VIBRATION-FREE MOUNTINGS WITH AUXILIARY MASS

by J. A. HARINGX.

621.752

The use of sensitive instruments such as balances, galvanometers and microscopes is often made difficult or even impossible by vibrations in the surroundings. In order to reduce the amplitudes of these forced vibrations the instruments can be placed upon sufficiently weak springs. Then, however, a damping must be introduced in order to stop the free vibrations of the system after a slight impulse or an initial displacement. The most obvious manner of applying this damping, namely between the apparatus and the foundation upon which it is placed, is indeed favourable for the rapid decay of the free vibrations, but it promotes the forced vibrations. A better method consists in introducing the damping between the apparatus and an auxiliary mass attached to it by means of springs. The features of this system and the choice of the different parameters (masses, rigidities of the springs, damping) are discussed in this article for the one-dimensional case.

Various systems of vibration-free mountings

When sensitive instruments such as balances, galvanometers, microscopes and dial gauges are used difficulties are often experienced due to vibrations transmitted to the apparatus through the floors, walls and tables. In such a case an arrangement will be needed in which the transmission of the vibrations to the instrument in question is avoided. Very good results can be obtained by using a spring mounting, but it is quite impossible to construct in this way a support which completely prevents the transmission of vibrations. It is only possible to limit the amplitudes of the forced vibrations of the measuring instrument to such a degree that these vibrations no longer present any difficulty, so to that extent one may then say that the instrument is supported „vibration-free".

When a resilient layer is introduced between the instrument and its foundation, for instance a rubber cushion or a set of helical springs, a system is obtained like that shown diagrammatically in *fig. 1a*: a mass m joined by a spring with the rigidity c (*i.e.* the force per unit of elongation of the spring) to a foundation which vibrates in a vertical direction with an angular frequency ω and an amplitude a_0.

This system behaves as follows. The mass m vibrates at the same frequency [1]) ω with the foundation, but with an amplitude a which depends very much on the frequency. From the differential equation for the motion of the mass it may be derived that

$$\frac{a}{a_0} = \left| \frac{c}{c - m\omega^2} \right| .$$

This "frequency characteristic" is shown in

[1]) For the sake of brevity we use the term "frequency" here meaning (unless otherwise stated) the angular frequency $\omega = 2\pi$ times the frequency.

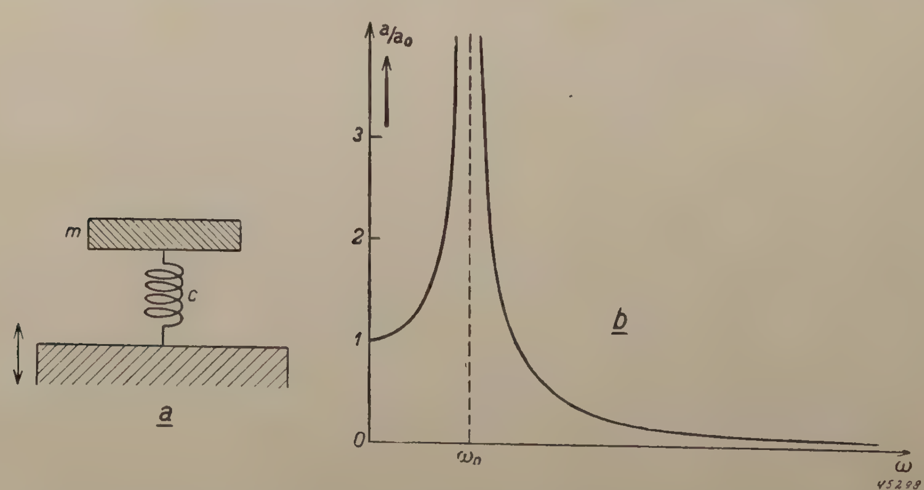


Fig. 1. Diagram (a) and frequency characteristic (b) of an undamped, vibration-free system. At frequencies ω which lie far enough above the resonant frequency ω_0 the amplitudes of the forced vibrations of the mass m caused by the vibration motion of the foundation are very much reduced. The amplitude ratio is: $\frac{a}{a_0} = \left| \frac{c}{c - m\omega^2} \right|$.

fig. 1*b*. In the neighbourhood of the frequency

$$\omega_0 = \sqrt{\frac{c}{m}},$$

the so-called resonant frequency of the system, the mass takes on very large amplitudes which are much larger than a_0. On the other hand at frequencies far enough above ω the amplitude becomes smaller than a_0. At very high frequencies it even gradually approaches the zero, changing in inverse proportion to the square of the frequency:

$$\frac{a}{a_0} \approx \frac{c}{m\omega^2} = \left(\frac{\omega_0}{\omega}\right)^2 \quad . \; . \; . \; . \; . \quad (1)$$

If the spring (c) is sufficiently weak and the mass (m) large enough to make the resonant frequency ω at least 3 to 5 times as low as the lowest frequency occurring in the interfering vibrations of the foundation, the instrument will take up these vibrations only to a very slight extent.

Besides the limitation of the amplitudes of the forced vibrations in the frequency region of the permanently occurring vibrations, it is also desired that after an impulse or initial displacement, caused either accidentally or by the operation of the instrument, the free vibrations of the system should come to rest as quickly as possible. Once it has taken up vibration energy the system according to fig. 1*a* continues to move and, theoretically, it takes an infinitely long time before the free vibration dies out. In practical systems, however, the vibration energy will gradually disappear, because there is always a certain damping even if it is only the internal damping of the material of the springs or the damping due to air resistance. It is clear that the stronger this damping the sooner the system will come to rest. This naturally suggests the introduction of an extra damping in the manner shown in *fig. 2a*. The damping force is assumed to be proportional to the relative velocity of the mass with respect to the foundation (viscous damping) and the proportionality factor is called k. As the free vibration dies out the amplitude then decreases as a function of time t proportional to $e^{-kt/2m}$. Such an arrangement, however, behaves entirely different from the first one, not only as far as the decay of the free vibrations is concerned but also as regards the forced vibrations. This is clearly shown by the frequency characteristics drawn in fig. 2*b* for different values of the damping k. The resonance peak is more or less reduced, but at the same time, at high frequencies, the decrease is much more gradual than in fig. 1*b*. It is found that in this case the amplitude ratio at high frequencies is inversely proportional to the frequency:

$$\frac{a}{a_0} \approx \frac{k}{m\omega}.$$

Thus the larger the damping k the more the resonance peak is cut down and the quicker the system comes to rest after an impulse, but at high frequencies the amplitudes are less effectively reduced;

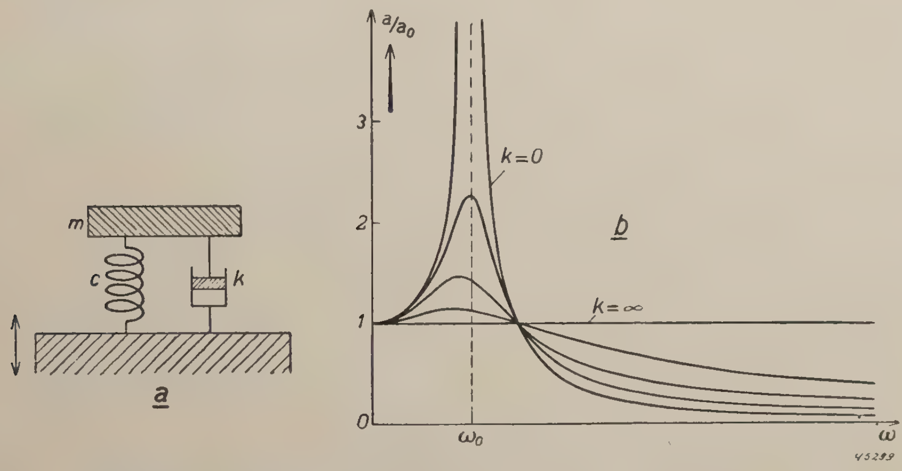


Fig. 2. Diagram (a) and frequency characteristic (b) of a vibration-free system with „relative" damping. With increasing damping (k) the resonance peak is gradually lowered and at the same time the free vibrations die out more quickly, but at high frequencies the amplitudes of the forced vibrations become larger. The amplitude ratio is given by the expression [2]):

$$\left(\frac{a}{a_0}\right)^2 = \frac{1 + q^2\bar{\omega}^2}{(\bar{\omega}^2 - 1)^2 + q^2\bar{\omega}^2}, \text{ where } \bar{\omega} = \frac{\omega}{\omega_0}, \; q = \frac{k}{m\omega_0}, \; \omega_0^2 = \frac{c}{m}.$$

[2]) See for example E. Lehr, Schwingungstechnik, Vol. II, pp 171 et seq. (J. Springer, Berlin 1934).

thus in order to limit sufficiently the amplitudes of the forced vibrations the resonant frequency will have to be placed farther below the interfering frequency region.

Since, however, in practice, the resonant frequency cannot be made arbitrarily low (if the mass should be limited the mounting would become too "weak"), an arrangement according to the principle of fig. 2*a* does not usually give satisfactory results: with low damping it takes too long for the free vibrations to die out, whereas with large damping amplitudes of the forced vibrations are not sufficiently limited.

The situation would be quite different if the damping were not introduced between the mass m and the vibrating foundation as in fig. 2*a* (relative damping), but between the mass and a fixed point in space, so-called absolute damping, see *fig. 3a*. The frequency characteristics of this system, drawn in fig. 3*b* for different values of the damping, show that in this case increased damping is advantageous in every respect. The resonance peak is very much flattened, while even at high frequencies, *i.e.* in the interfering frequency region of the foundation, has the effect of a rigid coupling. Thus, if in fig. 2*a* the damping increases, the mass m will finally be rigidly joined with the foundation and will therefore follow its movement completely ($a/a_0 = 1$). If, on the other hand, in fig. 3*a* the damping is increased, the mass is finally rigidly bound to a stationary point in space and thus remains completely at rest ($a/a_0 = 0$).

The case of absolute damping is of course only of theoretical interest. If it were actually possible to have a fixed point in space it would be advisable to make the instrument vibration-free by attaching it rigidly to the point in question directly, instead of mounting it with springs on the vibrating foundation.

Nevertheless, even without having a fixed point at our disposal, it is possible to imitate absolute damping to a certain extent by introducing the damping element (k) between the main mass m and an auxiliary mass M attached to the main mass by a spring (rigidity C). This arrangement is represented schematically in *fig. 4a*. At high frequencies the auxiliary mass M, due to its inertia, will in general have a tendency to remain at rest and in that way

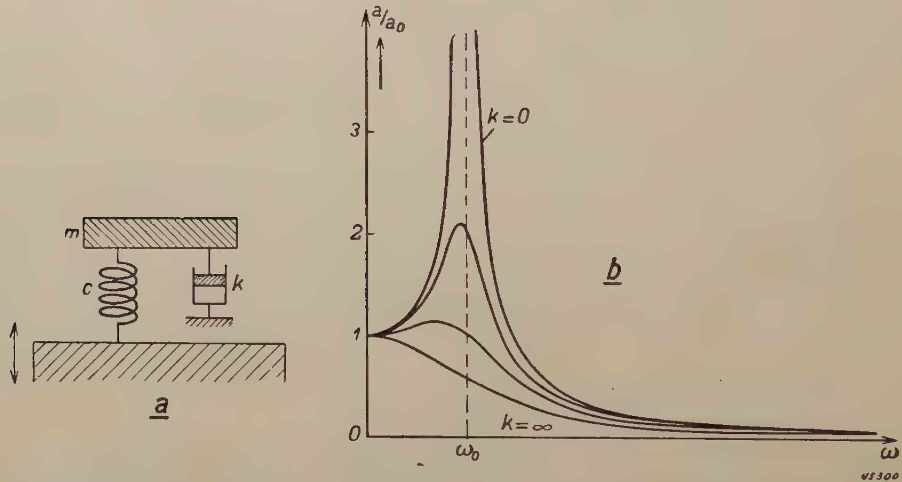


Fig. 3. Diagram (a) and frequency characteristic (b) of a vibration-free system with "absolute" damping. With increasing damping (k) not only is the resonance peak lowerd and the duration of the free vibrations shortened, but also the amplitudes of the forced vibrations are reduced. The amplitude ratio is given by the expression [3]):

$$\left(\frac{a}{a_0}\right)^2 = \frac{1}{(\bar\omega^2-1)^2+q^2\bar\omega^2}, \text{ where } \bar\omega = \frac{\omega}{\omega_0},\ q = \frac{k}{m\omega_0},\ \omega_0^2 = \frac{c}{m}.$$

the amplitude is always less than in the arrangement with no damping (fig. 1) and, as in that case, at sufficiently high frequencies its trend is again according to equation (1). The difference between the effect of this absolute damping and the relative damping indicated in fig. 2 is easily understood when one considers that the damping introduced between two points constitutes a hindrance to their relative movement and an infinitely large damping even will furnish an almost fixed — or at least a slightly vibrating — point of contact for the damping force. Such a conception is indeed confirmed theoretically and the arrangement is found to possess very favourable properties not only as regards the amplitudes of the forced vibrations but also in

[3]) See for example E. Lehr, Schwingungstechnik, Vol. II, pp 135 et seq. (J. Springer, Berlin 1934).

respect to the decay of the free vibrations. Several vibration-free mountings have already been constructed on this principle in the Philips factories.

We will now look more closely into the behaviour of a system according to fig. 4*a* and consider the choice of the various parameters (m, c, M, C, k), confirming our attention to the one-dimensional case.

upon the parameters of the system (masses and rigidities of the springs).

Finally, in the general case where the damping k has a value between 0 and ∞, the frequency characteristic always passes through the two points of intersection A and B of the two extreme curves first considered and for the rest lies entirely between those two curves, in general either with

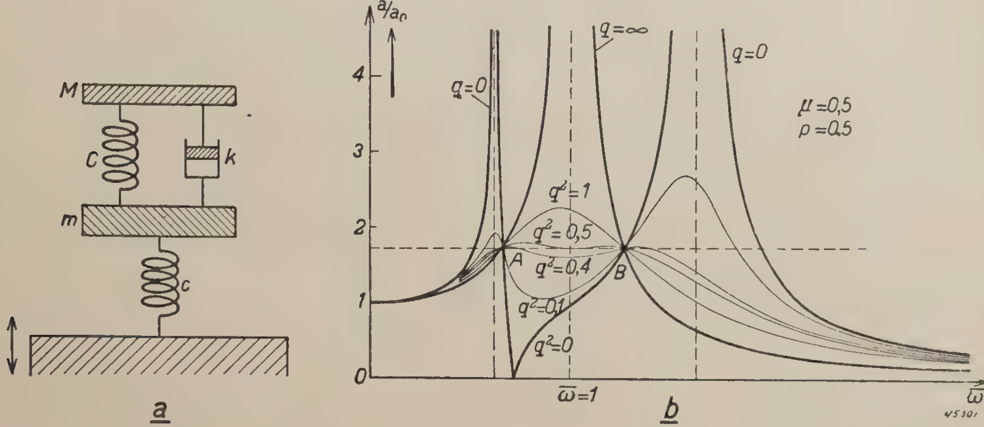


Fig. 4. Diagram (a) and frequency characteristic (b) of a vibration-free system with auxiliary mass M. The damping (k) is introduced between the main mass (m) and the auxiliary mass. The amplitude ratio is given by formula (3) on page 20. The figure here is drawn for the special case $\mu = 0.5$ and $p = 0.5$, where the points of intersection A and B of all the frequency characteristics lie at the same height.

The behaviour of the system with auxiliary mass in relation to the parameters

The frequency characteristic

In fig. 4*b* we have the frequency characteristic for a given case at different values of damping. We will begin to analyse this figure by taking the extreme case where the damping k is infinitely large. The auxiliary mass M is then rigidly bound to the main mass m; we then have in fact a system like that of fig. 1, with spring rigidity c and mass $m + M$. The frequency characteristic of this system is the fully drawn curve in *fig. 5* (identical with fig. 1*b*), having the resonance frequency

$$\omega_0 = \sqrt{c/(m + M)}.$$

Since $\bar{\omega} = \omega/\omega_0$ is taken as abscissa, resonance occurs at $\bar{\omega} = 1$.

If, on the other hand, $k = 0$, we have the familiar case of two coupled, undamped oscillators. Such a system with two degrees of freedom is in resonance at two different frequencies and therefore has a frequency characteristic like the dotted curve in fig. 5. The exact position of the resonant frequencies on either side of ω_0 ($\bar{\omega}$=1), as well as of the intermediate zero point P of the curve, depends

one maximum between A and B or with two maxima one on either side of A and B. See the family of curves in fig. 4*b*.

It will perhaps be interesting and useful to explain this in somewhat more detail with formulae, the derivation of which may be briefly outlined [4]). When considering the forces acting on the main and auxiliary masses, we get for the motion of the two masses two coupled linear differential equations of the second order, from which by elimination of the coordinate of the auxiliary mass a linear differential equation of the fourth order is obtained for the coordinate x of the main mass. If now the foundation vibrates according to $a_0 \sin \omega t$, the main mass will vibrate at the same frequency but generally with a different amplitude and phase: $x = a \sin(\omega t + \varphi)$. By substituting this in the differential equation, an equation can be derived for the frequency characteristic, namely for the ratio a/a_0 as a function of ω and of the parameters m, c, M, C, k. These quantities, however, are

[4]) Cf. J. P. den Hartog, Vibrations et mouvements vibratoires, p. 99 (Ed. Dunod, Paris 1936); E. Hahnkamm, Die Dämpfung von Fundamentschwingungen bei veränderlicher Erregerfrequenz, Ing. Arch. 4, 192, 1933; L. Geislinger, Theorie des Resonanzschwingungsdämpfers, Ing. Arch. 5, 146, 1934.

found to occur in the result only in certain combinations, so that the result can be written much more simply and be made more comprehensible by introducing the following (dimensionless) quantities:

$$\left.\begin{aligned} \bar{\omega} &= \frac{\omega}{\omega_0} = \omega \sqrt{\frac{m+M}{c}}, \\ p &= \frac{C}{c}\,\frac{m+M}{M}, \\ q &= \frac{k}{M}\sqrt{\frac{m+M}{c}}, \\ \mu &= \frac{m}{m+M}, \end{aligned}\right\} \quad \ldots \quad (2)$$

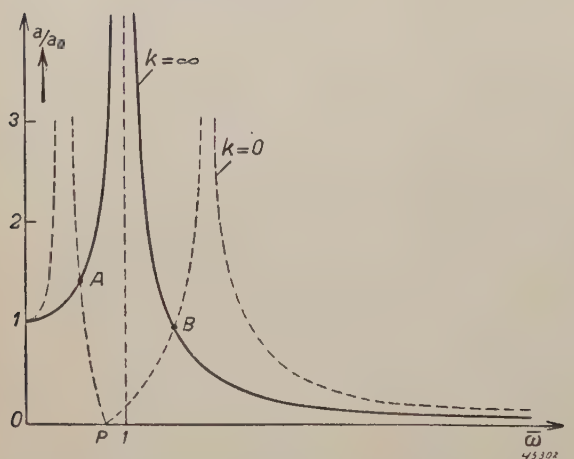


Fig. 5. Frequency characteristic of the system with auxiliary mass (fig. 4a) in the two extreme cases of damping $k = \infty$ and $k = 0$. In the first case main mass and auxiliary mass are rigidly connected and we have a system like that in fig. 1a (fully drawn curve), while in the second case we have a combination of two coupled (undamped) vibration systems. There are then two resonant frequencies (dotted curves) corresponding to the two degrees of freedom; $\bar{\omega} = \omega/\omega_0$ is the reduced frequency.

The quantity p will be called the rigidity parameter, q the damping parameter and μ the mass parameter. The formula for the frequency characteristic now becomes

$$\left(\frac{a}{a_0}\right)^2 = \frac{(\bar{\omega}^2 - p)^2 + q^2\bar{\omega}^2}{[\mu\bar{\omega}^4 - (1+p)\,\bar{\omega}^2 + p]^2 + q^2\bar{\omega}^2(\bar{\omega}^2 - 1)^2}. \quad (3)$$

In the case of infinitely large damping, thus $q = \infty$, the formula becomes

$$\left(\frac{a}{a_0}\right)_{q=\infty} = \left|\frac{1}{\bar{\omega}^2 - 1}\right| \quad \ldots \quad (4)$$

This is the equation for the fully drawn curve in fig. 5. In the case without damping, thus $q = 0$, the formula becomes

$$\left(\frac{a}{a_0}\right)_{q=0} = \left|\frac{\bar{\omega}^2 - p}{\mu\bar{\omega}^4 - (1+p)\,\bar{\omega}^2 + p}\right|,$$

which is the equation for the dotted curve in fig. 5.

If for the sake of simplicity we now call the functions of ω occurring in the four terms of numerator and denominator of (3) α, β, γ, δ, we may then write for (3):

$$\left(\frac{a}{a_0}\right)^2 = \frac{\alpha + q^2\beta}{\gamma + q^2\delta} = \frac{\alpha\,(1 + q^2\cdot\beta/\alpha)}{\gamma\,(1 + q^2\cdot\delta/\gamma)}$$

and it is immediately clear that with the condition that

$$\frac{\beta}{\alpha} = \frac{\delta}{\gamma} \quad \ldots\ldots\ldots \quad (5)$$

the quotient $(a/a_0)^2$ becomes equal to α/γ and thus independent of q. At those values of ω for which (5) is satisfied, therefore, the curves have the same ordinate for all values of q, including $q = 0$ and $q = \infty$, from which it follows, as already stated, that all the curves pass through the two points of intersection A and B of the fully drawn and the dotted curves in fig. 5.

For the abscissae $\bar{\omega}_A$ and $\bar{\omega}_B$ of these points of intersection, which we shall presently need, we can derive the following equation by substituting the four functions $\alpha \ldots \delta$ in (5):

$$\bar{\omega}^4 - 2\,\frac{1+p}{1+\mu}\,\bar{\omega}^2 + 2\,\frac{p}{1+\mu} = 0,$$

thus:

$$(\bar{\omega}^2)_{A,B} = \frac{1}{1+\mu}\,(1 + p \pm \sqrt{p^2 - 2p\,\mu + 1}), \quad (6)$$

where

$$\bar{\omega}_A < 1 \text{ and } \bar{\omega}_B > 1.$$

For the ordinates of the points of intersection one then finds according to (4)

$$\left(\frac{a}{a_0}\right)_A = \frac{1}{1 - \bar{\omega}_A^2} \text{ and } \left(\frac{a}{a_0}\right)_B = \frac{1}{\bar{\omega}_B^2 - 1}. \quad (7)$$

Behaviour at high frequencies

From the general shape of the curves in fig. 4b it may be seen that also for this type of system the resonance region must in any case lie far below the interfering frequency region of the vibrations of the foundation. The degree to which the amplitudes of the forced vibrations of the main mass are then restricted can easily be deduced from equation (3). For sufficiently high frequencies it becomes

$$\left|\frac{a}{a_0}\right| \approx \frac{1}{\mu\bar{\omega}^2} = \frac{c}{m\omega^2} \quad \ldots \quad (8)$$

When we compare (8) with (1) we see that with the same values of c and m we obtain the same

favourable behaviour as in the elementary arrangement entirely without damping (fig. 1) or as in that with absolute damping (fig. 3). The presence of the auxiliary mass M therefore apparently plays no part here.

Such behaviour has already been presumed quantitatively from the tendency of the auxiliary mass M to remain stationary at high frequencies.

Actually, of course, M does move slightly, and it may be deduced that the amplitude a of this movement at high frequencies is determined by $|a'/a/| \approx kc/mM\omega^3$. The ratio a'/a thus decreases more rapidly with increasing ω than the amplitude ratio for the main mass according to formula (8).

The optimum choice of parameters

In the practical realisation of a vibration-free mounting according to fig. 4 the question will naturally arise as to how stiff the springs must be made, and how heavy the masses and how strong the damping will have to be. The behaviour at high frequencies, equation (8), gives the indication already known, that the resonance region of the system must lie at the lowest frequencies possible. As to the choice of C, M and k this does not help us at all, since according to equation (8) these parameters do not affect the amplitudes of the forced vibrations at high frequencies. The choice of these parameters will, however, be decisive for the behaviour at lower frequencies and for the decay of the free vibrations of the system after an accidental impulse or initial displacement.

During the free vibrations the motion of the system is in general composed of two vibrations with the two "resonant frequencies", the amplitude of each of these vibrations decreasing exponentially with the time: $e^{-a_1 t}$ and $e^{-a_2 t}$ respectively, while the relation between the initial amplitudes of the two vibrations depends upon the initial conditions (initial displacement or impulse). The rate of decay of the free vibration will thus depend not only on a_1 and a_2 but also on the accidental initial conditions. It is therefore impossible to speak of *the* rate of decay of the free vibrations, and even when the initial conditions are given the influence of the parameters of the system on the decay of the vibrations is still very difficult to ascertain. When, however, we assume that the limitation of the duration of the free vibrations runs more or less parallel with the decrease in the amplitudes of the system in the resonance region, thus in a manner similar to the behaviour of the systems with one mass (figs. 2 and 3), we only need to study the frequency characteristic at the lower frequencies. We might then state the condition that the highest peak in the frequency characteristic should be as low as possible, and from that requirement derive the optimum choice of the parameters.

Among the curves with different values of the damping parameter q there will be one which has its highest peak just at the highest of the two points A and B. When the position of A and B is known this is evidently the most favourable possibility, because then the ordinate of A or B is never exceeded. The value of q corresponding to this curve is, it is true, still unknown, but that does not affect our argument. We now study the position of A and B with reference to fig. 5. The fully drawn curve ($q = \infty$) to which equation (4) applies is entirely fixed if the resonant frequency ω_0 is taken as given. The points A and B will then always lie on this curve, and from formula (6) it can be proved that when we vary the parameter p both points are displaced to the right or to the left. From fig. 5 it may be seen that one point therefore always rises as the other falls, and *vice versa*. The highest peak of the frequency characteristic can then also be considerably lowered with respect to fig. 5 by giving p a value such that the points A and B lie at the same height as is the case in fig. 4*b*. Although the optimum frequency characteristic then exhibits two maxima lying at the same height but coinciding neither with A nor with B, the differences are so extremely small that in this way a good approximation is attained. From (7) it follows that A and B lie at the same height when the following condition is satisfied:

$$\bar{\omega}_A^2 + \bar{\omega}_B^2 = 2.$$

If we substitute here expression (6) for $\bar{\omega}_A$ and $\bar{\omega}_B$ we obtain as "optimum" value for the rigidity parameter

$$p_{\text{opt}} = \mu.$$

For the corresponding maximum amplitude ratio we find

$$\left(\frac{a}{a_0}\right)^2_{\text{opt}} \approx \left(\frac{a}{a_0}\right)^2_{A,B} = \frac{1+\mu}{1-\mu} \quad . \quad . \quad . \quad . \quad (9)$$

On the basis of the calculations of Collatz [5]) it can further be shown that the optimum value of the damping parameter can be taken with a very close approximation to be:

$$q_{\text{opt}} \approx \sqrt{1{,}5\,\mu\,(1-\mu)}.$$

---

[5]) L. Collatz, Über den günstigsten Wert der Kopplungskonstanten bei reibungsgekoppelten Systemen, Ing. Arch. **10**, 269, 1939.

Important practical case: mass parameter $\mu = 0.5$

After the foregoing the problem of the choice of the parameters is reduced to the choice of the mass parameter $\mu = m/(m+M)$. Since we try to keep the frequency characteristic as low as possible we must try to make the largest ordinate now occurring as small as possible. According to equation (9) the smallest possible value of μ is desired, *i.e.* in our vibration-free mounting the auxiliary mass M would have to be large compared with the main mass m. In practice, however, one is not likely to make M larger than the main mass. If we therefore assume that the two masses are chosen of equal size, thus $\mu = 0.5$, we obtain:

$$p_{opt} = \mu = 0.5,$$

$$q_{opt} \approx \sqrt{1.5\,\mu\,(1-\mu)} = 0.612.$$

Having regard to equation (2), it also appears that the following relations must be satisfied:

$$C/c = \mu(1-\mu) = 0.25,$$

$$\frac{k}{\sqrt{mc}} \approx (1-\mu)\sqrt{1.5\,(1-\mu)} = 0.433.$$

Finally, the maximum amplitude ratio according to equation (9) amounts to only

$$\left(\frac{a}{a_0}\right)_{opt} \approx \sqrt{\frac{1+\mu}{1-\mu}} = 1.728.$$

When $p = \mu = 0.5$ we find exactly [6])

$$q_{opt} = 0.624 \text{ and } (a/a_0)_{opt} = 1.746.$$

The approximations given are thus found to agree very well with the exact values. As a matter of fact even if we make the damping parameter much too large or too small, (a/a_0) still varies only slightly; between $q = 0.4$ and $q = 0.95$ the increase of the amplitude ratio with respect to its optimum value (1.75) amounts at the most to 25%. The same is true for the choice of p. If we keep to $\mu = 0.5$ and take in each case the best value of q, a variation of p between 0.3 and 0.7 results in a maximum rise of 25% in $(a/a_0)_{max}$.

When we compare these optimum results of our mounting with auxiliary mass with the vibration-free mounting without auxiliary mass, we assume the total mass $m + M$ or m and the rigidity of the spring c to be given, since in the practical construction a certain weight will in general be available and it is, moreover, required that the resilient attachment of the main mass should possess a certain degree of rigidity. The resonant frequencies

$$\omega_0 = \sqrt{c/(m+M)} \text{ resp. } \omega_0 = \sqrt{c/m}$$

are then automatically equal, so that for each ω the reduced frequencies $\overline{\omega} = \omega/\omega_0$ have the same value in both cases. In *fig. 6* the following curves are given for the sake of comparison.

1. The frequency characteristic of the undamped system according to fig. 1*a*;
2. that of the system with auxiliary mass according to fig. 4*a* with the same rigidity c and the same total mass divided into two equal parts: thus μ is 0.5 and further $p = 0.5$; $q = 0.62$;
3. that of the system according to fig. 2*a* with relative damping, with the same rigidity and mass and a damping ($q = k/m\omega_0 = 0.74$) such that the maximum of the frequency characteristic lies just as low as that of curve *2*.

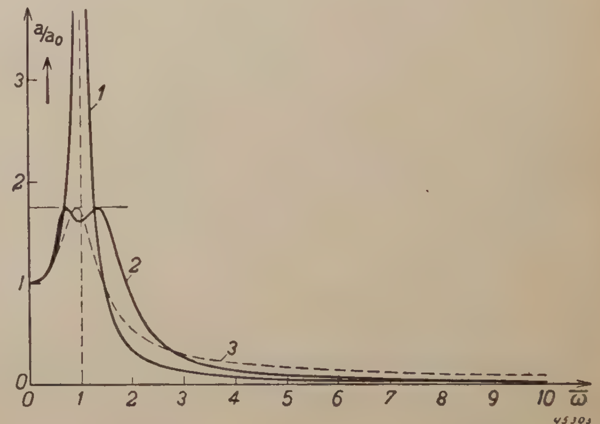


Fig. 6. Frequency characteristics of an undamped vibration system (curve *1*), a system with auxiliary mass and "optimum" parameters chosen for $\mu = 0.5$ (curve 2), and a system with relative damping (curve 3). It is assumed that the total mass and also the rigidity of the spring c is the same in all three systems. Further for curve *3* the damping was so chosen that the maximum amplitude ratio is the same as in case 2.

It is clearly seen how in case (3) the trend of the frequency characteristic is much less favourable at high frequencies. [6]) Moreover, the last mentioned system has the undesirable property that an accidental increase in the damping coefficient k (for instance with viscous damping due to a fall in temperature) causes the mass to vibrate with a proportionally larger amplitude, with the result that the system, intended as a vibration-free mounting, becomes much less effective.

On the other hand, if an auxiliary mass is applied the damping does not — or at least not perceptibly — affect the amplitudes of the forced vibrations at high frequencies, while, as we have seen, the maximum amplitude ratio in the resonance region reacts to a change in damping only to a slight extent.

---

[6]) Because $a/a_0 \approx km/\omega = q/\overline{\omega}$ instead of $a/a_0 = 1/\overline{\omega}^2$ and $1/\mu\overline{\omega}^2$ respectively.

Decay of the free vibrations

In order to avoid the difficulties encountered in studying the effect of the parameters on the behaviour of the system with auxiliary mass with respect to the decay of the free vibrations after an impulse or an initial displacement, we have adopted the simple point of view that the duration of the free vibrations would be restricted parallel with the reduction of the amplitudes at resonance, and consequently we aimed at the lowest possible maximum of the frequency characteristic. However, once all the parameters have been chosen, the rate of decay of the free vibrations can be determined exactly. For the case where $m = M$ with the corresponding "optimum" values $p = 0.5$ and $a = 0.62$ we find that after a certain initial displacement the free vibration of our main mass diminishes in the manner shown in *fig. 7*.

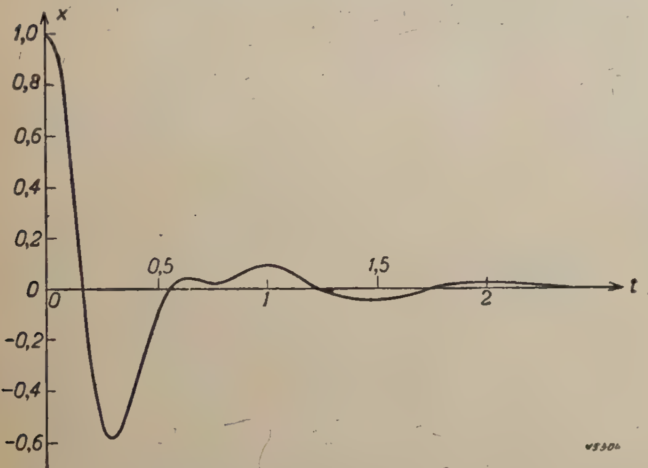


Fig. 7. The decay of the free vibration after a given initial displacement for a system according to fig. 4*a*, with $\mu = 0.5$ and the "optimum" parameters $p = 0.5$ and $q = 0.62$.

After the first period already the deviation has fallen to one-tenth of the initial displacement, while after two periods only $1^1/_2\%$ of the initial amplitude remains. This very favourable behaviour shows that our choice of parameters was a good one, although of course we may not assume that we have found the most favourable conditions with respect to the rate of decay of the free vibrations [7]).

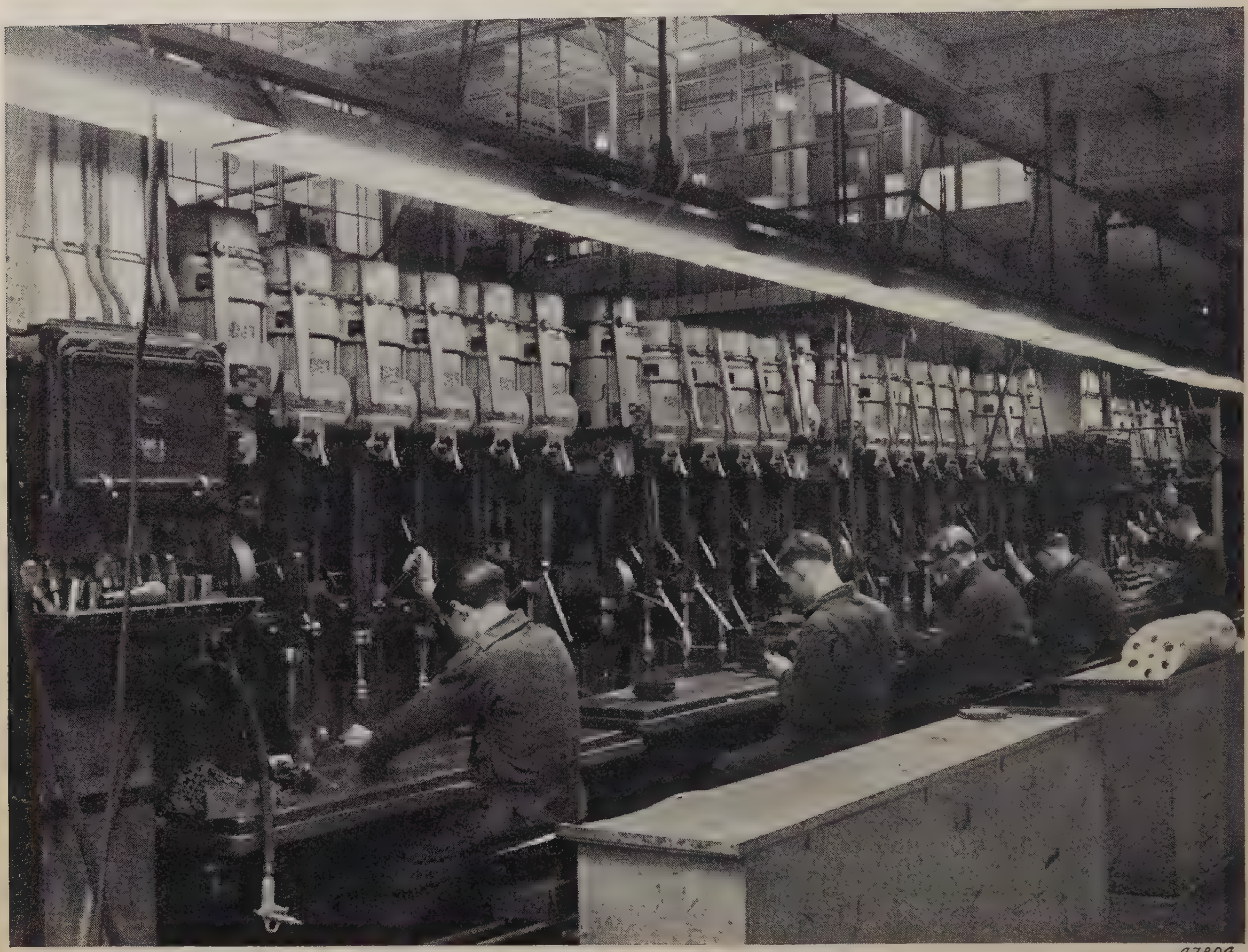
If, under the same initial conditions, we likewise investigate the free vibration of the elementary system with relative damping and with the same maximum in the frequency characteristic (case 3), we find that in this case the system comes to rest about one and a half times as quickly.

When it comes to putting the theory here developed into practical application for the construction of vibration-free mountings, one is faced immediately with the fact that as a rule the foundation of an apparatus may vibrate in different directions. We hope to discuss this more general case in a following article in this periodical, where at the same time we shall deal with the practical construction in more detail.

---

[7]) Closer consideration shows, for example, that in general at $p = \mu$ and $q \approx \sqrt{4\mu\,(1\text{-}\mu)}$ the system comes to rest somewhat more quickly than with our "optimum" values of the parameters $p = \mu$ and $q = \sqrt{1.5\,\mu\,(1\text{-}\mu)}$. These considerations, which would take us too far afield here, will be published elesewhere.

FACTORY LICHTING WITH GAS-DISCHARGE LAMPS



47804

Several articles have appeared in this journal from time to time dealing with the development, properties and applications of gas-discharge lamps.

The photograph reproduced here shows the lighting of 5 groups each of 5 drilling machines with gas-discharge lamps, type TL 100, mounted in metallic reflectors suspended $1^1/_2$ meters (abt. 4′10″) above the working plane. The intensity of light on the working plane is 100-150 lux, which is amply sufficient. This replaces the lighting system with movable reflectors fitted with incandescent lamps, and it is much more satisfactory.

STABILISED AMPLIFIERS

by J. J. ZAALBERG van ZELST.

621.394.645.3

At a given frequency the degree of amplification of an amplifier depends upon the properties of the valves used (particularly their slope) and the impedances of the other elements in the circuit. Assuming that the latter are fairly constant (given the right choice of material and proper construction, it is then a matter of designing the circuits in such a way as to be dependent as little as possible upon the properties of the valves, which may vary according to the amplitude of the anode voltages, the temperature, the contact potentials, etc., or when a valve is replaced. This is of great importance, for instance, when taking measurements.
In this article two groups of circuits are discussed (each of which may be divided into two sub-groups) which tend to provide for a high degree of constancy in the amplification. Some of these circuiting schemes need very few extra parts. Sometimes there is the additional advantage of reduced distortion.
In many cases combinations of the various methods are possible, resulting in an exceptionally constant amplification.

Introduction

In many cases, for instance for measuring purposes, it is desired to have an amplifier that does not change with time, even though variations may occur in such factors as anode voltages, contact potentials, temperature, etc. which may affect the amplification directly or indirectly. This demand should not be taken too literally; if the factors referred to remain within certain reasonable limits — limits which in practice are exceeded only in exceptional cases — it is sufficient if the amplification is also kept constant within certain limits narrow enough for the deviations from the nominal value to be negligible for the purpose in view. Here in this article we will deal with the principles that count in the designing of such an amplifier as this, confining ourselves to those cases where the amplifying action is obtained through the alternating voltage to be amplified between two electrodes of a valve (*e.g.* cathode and control grid) resulting in a current of the same frequency to another electrode of the valve (*e.g.* the anode). We will consider only the influence that the variations of the valve properties have upon the amplification, because the changes taking place in the other elements of a circuit can be kept within certain narrow limits by suitable construction or choice of material.

In the arrangement as found in an amplifier there is for each valve or each set of valves a certain relation between the alternating current generated and the alternating voltage applied, a relation which depends somewhat upon the amplitude but with the usual order of magnitude only to a small extent. For the sake of brevity this relation will here be termed the slope (symbolised by S), generalising somewhat the usual meaning of the word. The actual amplification, *i.e.* the ratio of the output voltage to the input voltage, is proportional to that slope.

The methods for improving stability of the amplification may be divided into two groups. In the first group the valves retain their slope as given by circumstances, the improved stability being obtained by adding a compensating quantity to the input or output signal; this can be done outside the amplifier, so that nothing need be altered in the amplifier itself.

This group comprises:

Ia) feedback: part of the output signal is fed back to the input circuit and amplified with it;

Ib) the input signal is equated with a part of the output signal and the difference, after being amplified in an extra amplifier, is added to the output signal.

In the second group of methods the operation of the valves is arranged in such a way that the slope remains constant, using a control voltage derived from an auxiliary alternating voltage of a frequency that does not cause any interference. There are two possibilities, according to the origin of the auxiliary voltage:

IIa) where it is extraneous to the amplifier;

IIb) where it is generated in the amplifier itself.

Each of these methods will now be considered in turn.

Ia) Adding a compensating quantity to the input signal (feedback)

Much has already been written, also in this periodical [1]), about feedback and in particular

[1]) Philips Techn. Review 1, 268, 1936 and 2, 289. 1937,

about the special form of feedback termed **negative feedback**, which we have mainly in mind here. Nevertheless, it is well to recall briefly the principle of this system.

The output current I_a of an amplifier A (*fig. 1*) or the output voltage as the case may be, or a part

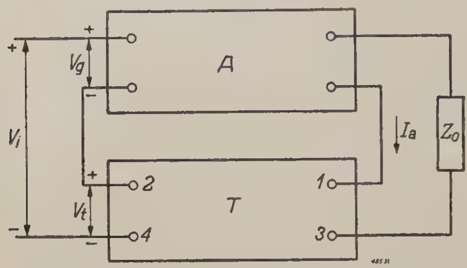


Fig. 1. The output a.c. current I_a from an amplifier A is conducted to the input terminals *1-3* of a feedback system T. At the output terminals *2-4* there consequently arises a voltage V_t which together with the input signal V_i forms the voltage V_g supplied to A. The + and — signs at the terminals indicate in what direction the a.c. voltages V_i, V_t and V_g are positive. Z_o is an external impedance.

of it, is conducted to a network T called the feedback circuit. The output voltage V_t of this circuit, together with the signal V_i to be amplified, forms the input voltage V_g of the amplifier. One speaks of **negative feedback** when the phase difference between V_i and V_t is such that the amplitude of V_g is less than that of V_i.

The amplifier A may be characterised by the slope S:

$$I_a = SV_g, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

the feedback circuit T by a **transfer impedance** Z:

$$V_t = ZI_a = (R + jX)\, I_a \quad . \quad . \quad . \quad . \quad (2)$$

The terminals *1* and *2* of the circuit T may be coincident; equally so the terminals *3* and *4*. In that event Z is simply the impedance connected between the two points *1-2* and *3-4*. In more complicated cases the network as shown in fig. 1 has four poles and Z then indicates the ratio of the output voltage to the input current of this four-polar system; one then speaks of a **transfer impedance**.

Finally, with the positive direction of the voltages indicated in fig. 1 we have the relation

$$V_g = V_i - V_t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

In the absence of feedback ($Z = 0$, thus $V_t = 0$) the slope S' of the whole system is identical with that of the amplifier itself:

$$S' = \frac{I_a}{V_i} = \frac{I_a}{V_g} = S.$$

Where feedback is applied we find however:

$$S' = \frac{I_a}{V_i} = \frac{SV_g}{V_i} =$$

$$= S \cdot \frac{V_i - V_t}{V_i} = S\left(1 - \frac{ZI_a}{V_i}\right) = S\,(1 - ZS')$$

or

$$S' = \frac{1}{Z + \frac{1}{S}} = \frac{1}{\frac{1}{S} + R + jX},$$

hence

$$|S'| = S_{eff} = \frac{1}{\sqrt{\left(\frac{1}{S} + R\right)^2 + X^2}}, \quad . \quad . \quad . \quad (4)$$

where $|S'|$ is denoted by S_{eff}.

The question is now how R and X can best be chosen so that certain variations of S have the minimum influence upon S_{eff} without the latter becoming appreciably smaller than the original slope S. It is to be borne in mind that the components R and X of the transfer impedance Z may be negative, whilst the four-polar system T need not necessarily have any negative resistance, self-inductances or capacities (an example of such a case will be given presently). Consequently the aim will be to give R such a negative value as will just compensate the mean value of $1/S$. Supposing that through some cause or other S fluctuates between the limits S_{min} and S_{max}, one will then choose

$$R = -\tfrac{1}{2}\left(\frac{1}{S_{\max}} + \frac{1}{S_{\min}}\right) \quad . \quad . \quad . \quad (5)$$

By this means the first term in the denominator of equation (4) is caused to disappear as far as possible, with the result that 1) variations of S — which factor occurs only in this term — have the least possible influence and 2) for a given value of X the denominator of (4) is reduced to the lowest possible value and consequently the amplification becomes as high as possible.

By substituting (5) in (4) one finds that the effective slope will lie between the limits

$$S_{eff\,\max} = \frac{1}{|X|}$$

and

$$S_{eff\,\min} = \frac{1}{\sqrt{\frac{1}{4}\left(\frac{1}{S_{\min}} - \frac{1}{S_{\max}}\right)^2 + X^2}} \quad . \quad (6)$$

Such a variation can be made relatively as small as one desires by choosing $|X|$ only just large enough; as $|X|$ is increased, however, so the amplification is reduced, and one must therefore find the

compromise best suitable for each particular case. Given a sufficiently low value of $|X|$, S_{eff} becomes greater than the original S. According as S_{eff} is greater or smaller than S, so one gets positive or negative feedback respectively. As may be calculated with the aid of eq. (6), when applying negative feedback the sacrifice in amplification is accompanied by a large gain in stability; with a weak positive feedback there is much less gain and a strong positive feedback even results in a loss of stability. This will be illustrated by a concrete example.

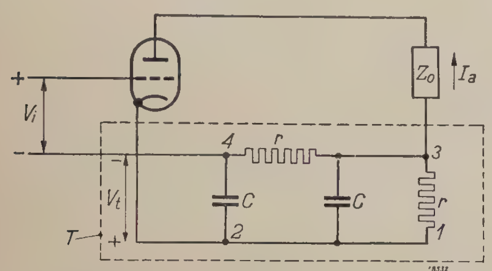


Fig. 2. Example (see footnote [2]) of an amplifier with feedback system T consisting of two resistances r and two capacitors C. The numbering *1-4* of the terminals of T corresponds to that of fig. 1.

Fig. 2 shows the circuit diagram [2]) of an amplifier where the feedback circuit consists of a combination of two resistances and two capacitors. For the sake of simplicity it is assumed that both resistances have the value r and both the capacitors the value C, though this is by no means essential.

As a simple calculation will show, for an angular frequency ω we get for this circuit

$$R = \frac{1-p^2}{1+7\,p^2+p^4}\,r, \quad X = \frac{-3p}{1+7\,p^2+p^4}\,r, \quad (7)$$

where $p = \omega Cr$.

If, for example, an average slope of 5 mA/V is subject to variations of + and — 10%, we find from eq. (5) for R the value —202 ohms. From (7) it appears that to get negative values of R it is necessary that p should be greater than unity. If one chooses for instance $p = 2$ and the frequency of the signal to be amplified is say 1000 c/sec, it then follows from (7) that $r = 3030$ ohms, $C = 0.105$ μF, $X = -404$ ohms, S_{eff} is then 2.5 mA/V ± 0.064%, so that by sacrificing only a factor of 2 in slope an enormous gain is obtained in stability.

It would also have been possible to leave the slope unaltered, by choosing $X = 1/S$, thus $X = -200$ ohms. From eq. (7) it follows that in that case we have to take $r = 4000$ ohms and $C = 0.131$ μF. Eq. (6) shows that the result is then $S_{eff} = 5$ mA/V ± 0.25%, which for many purposes is still quite satisfactory.

[2]) For the sake of clarity the sources of grid and anode voltage have been omitted in figs. 2 and 8.

Any gain in amplification can only be realised at the cost of gain in stability. In the example just given, for instance, the average slope could be increased say by a factor of 5, thus to 25 mA/V, by choosing $X = -38.2$ ohms, taking $r = 54\,500$ ohms and $C = 0.047$ μF (again for 1000 c/sec), but with the assumed 10% variation in S there would still be a fluctuation of 6.3% in S_{eff}, so that in this respect there is no improvement worth mentioning. If a still stronger positive feedback were to be applied S_{eff} would in fact fluctuate much more than S.

In the designing of the feedback circuit care is to be taken to avoid oscillation, which would be undesirable. If there is a freqeuncy for which the value Z assumes the proportions of $1/S$ then the amplifier will start oscillating in that frequency.

This can be explained as follows: Suppose that the output of the feedback system T and the input of the amplifier A (*fig. 3*) are disconnected from each other for a moment while there is no signal to be amplified. Upon applying a voltage V_{t1} to the amplifier an anode current $I_a = SV_{t1}$ is generated, which in turn supplies a voltage $V_{t2} = ZI_a = ZSV_{t1}$ to the output of the system. The condition for $V_{t2} = V_{t1}$ is therefore

$$ZS = -1 \quad \ldots \ldots \ldots \quad (8)$$

If that condition is satisfied and V_{t1} were of that frequency, then upon T and A being connected again the situation would remain as it was; in other words the amplifier would oscillate.

With the feedback system of fig. 2 there is no risk of oscillation, for the limit of $ZS = -1$ for oscillation can only be reached when Z is the real value (assuming that S is a real value, which is usually the case), thus when $X = 0$, which according to eq. (7) is only the case for $\omega = 0$ and $\omega = \infty$ (moreover with $\omega = 0$ according to eq. (7) R is

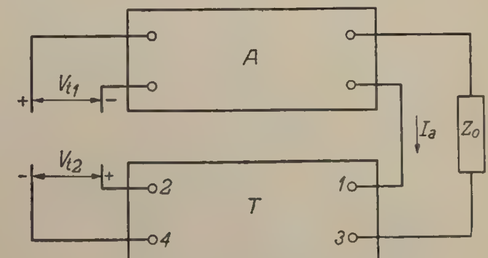


Fig. 3. When for a certain frequency the transfer impedance $Z = V_{t2}/I_a$ of the feedback system T assumes the value $-1/S$ (eq. (8)) then the voltage supplied by T is just identical with the voltage required at the input of the amplifier A to produce V_{t2}. When the output of T is connected to the input of A the amplifier begins to oscillate in that frequency.

positive, *viz.* equal to r, so that with this solution equation (8) $ZS = -1$ is not satisfied.

Finally it is to be remembered that when applying negative feedback the non-linear distortion may be reduced [3]).

Summarizing, we come to the conclusion that this method of feedback answers the purpose and is simple, with little or no sacrifice of amplification, and that it may lead to reduced distortion. It is to be borne in mind, however, that both the amplification and the improvement of stability are functions of the frequency, so that as a rule this method is less suitable if the signal to be amplified covers a wide frequency band. Furthermore, unless the feedback system is properly circuited there is a risk of troublesome oscillation.

Ib) Addition of a compensating quantity to the output signal

Briefly the principle of this method lies in the output voltage of an amplifier being reduced, by means of, for instance, a potentiometer or a transformer, in a ratio equal to the amplification required (which we shall call A_o), the fraction of the output voltage thus obtained being compared with the input voltage. If the amplification were exactly A_o then that fraction would be equal to the input voltage. Any difference is conducted to a separate amplifier having an amplification factor A_o. Combination of the output voltage of the two amplifiers then produces a voltage exactly A_o times the input voltage.

When, instead of the output voltage, the anode current is taken as the output signal, then one proceeds as follows: with the help of a transfer impedance of the order of $1/S_o$ one derives from the anode current I_{a1} of the main amplifier with the desired slope S_c a voltage that is equated with the input voltage. Any difference between these two voltages results in an anode current I_{a2} in the anode circuit of the auxiliary amplifier (slope S_o), and the sum of the two anode currents is exactly S_o times the input voltage.

Schematically this could be represented as follows: let $S_o + \Delta S$ be the actual slope of the main amplifier, then

1) the input voltage V_i, from which we start, supplies

2) in the main amplifier an anode current

$$I_{a1} = (S_0 \pm \Delta S) \cdot V_i.$$

3) With the aid of a transfer impedance $1/S_c$ we derive from this a voltage

$$\frac{S_0 \pm \Delta S}{S_0} V_i = \left(1 \pm \frac{\Delta S}{S_0}\right) V_i$$

4) and compare this with the input voltage. The difference is

$$\mp \frac{\Delta S}{S_0} V_i.$$

5) This difference is conducted to the auxiliary amplifier with slope S_o, which therefore supplies the anode current

$$I_{a2} = \mp V_i \Delta S,$$

6) which, added to the anode current I_{a1} of the main amplifier, produces just the desired output current $S_0 V_i$.

It may be thought that this is only transferring the difficulty to the auxiliary channel, the amplification of which was assumed to be A_o and the slope S_o but which may, of course, likewise be subject to variations. It must not be forgotten, however, that the auxiliary amplifier only supplies a compensating current, so that any fluctuations that may occur in this will have very little effect upon the ultimate result. This will be made quite obvious from a numerical example.

Suppose that a signal of 1V is required to yield a current of 100 mA, but that the slope of the main channel happens to be only 95 mA/V. The output current of 95 mA is conducted through a transfer impedance of 10 ohms, so that at the output terminals of that impedance we get a voltage of 0.95 V. The difference between this and the original signal, 0.05 V, is conducted to the auxiliary amplifier. If the latter has a slope of exactly 100 mA/V it will produce just the current of 5 mA lacking at the output of the main amplifier, but if for instance that slope should be 90 or 110 mA/V the final result would be 95 + (90 or 110) · 0.05 = 99.5 or 100.5 mA. Consequently a deviation of 10% in the slope of the auxiliary amplifier results in an error of only 0.5% in the total amplification.

This can be achieved both for a narrow and for a broad frequency band, according to the dimensions of the coupling elements (in contrast to the first method, which does not lend itself so well for a broad bandwidth).

It is easily realised that with this second method it is also possible to counteract the non-linear distortion. For instance a peak cut off in the main amplifier through over-loading is supplemented from the auxiliary channel.

[3]) See the articles referred to in footnote [1]).

It will be equally obvious that if desired a second or third auxiliary channel can be employed to reach still greater accuracy. *Fig. 4* gives an example of a system with one auxiliary channel (for details see the text below the diagram), but all sorts of variations are possible, which it is not necessary to enter into here.

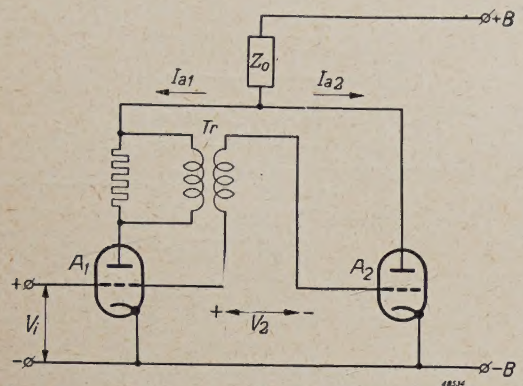


Fig. 4. The secondary voltage V_2 from the transformer *Tr* is just equal to the input voltage V_i when the slope of the main amplifier valve A_1 has the nominal value. Any deviation from that nominal slope causes a difference to arise between V_i and V_2, which is then conducted to the auxiliary amplifier valve A_2. The slope of A_2 is such that the anode current I_{a_2} just compensates the surplus of deficit in the anode current I_{a_1}, so that a constant current flows through the external impedance Z_o. $+B$ = positive, $-B$ = negative pole of the anode voltage source.

Summarizing briefly, the pros and cons of this method balance out as follows:

Advantages: no amplification loss in the main channel; greatly reduced distortion; no risk of oscillation; suitable for a broad frequency band.

Disadvantages: requires some additional circuiting elements, including at least one amplifying valve (the auxiliary channel, however, only need be dimensioned for a much smaller power than the main amplifier).

IIa) Controlling the slope with a separately generated auxiliary voltage

In this group of circuiting systems an auxiliary signal of a non-interfering frequency is supplied to the amplifier together with the main signal. On the output side the auxiliary signal is filtered out, rectified and smoothed, and the resultant d.c. voltage, after deduction of a fixed amount, is utilised as control voltage for the amplifying valve or valves.

Fig. 5 gives a diagram of an amplifier stage for high frequency. HF_i and HF_o are respectively the input and output terminals. An auxiliary voltage with low frequency is supplied to the transformer Tr_1. The voltage across the winding *sec 1* when rectified yields the d.c. voltage V_1. The low frequency a.c. voltage of the winding *sec 2* is amplified together with the h.f. signal and produces — *via* the tuned transformer Tr_2 and rectified by the diode D_2 — the d.c. voltage V_2.

The d.c. voltages V_1 and V_2 are both taken up in the control grid circuit of valve *A*, in such a way that V_1 works in a positive sense and V_2 in a negative sense. The ratio of the a.c. voltages from *sec 1* and *sec 2* and the transforming ratio of Tr_2 are of such dimensions that if the slope of *A* is of the right value V_1 and V_2 just compensate each other. If, however, the slope of *A* should decrease then V_2 drops, making the control grid voltage of *A* less negative, so that the valve operates in a part of the response curve where the slope is larger; with an increasing slope the reverse takes place. In this way any deviation from the normal slope is automatically corrected, at least in part.

Any variations in the amplitude of the l.f. input signal affect V_1 and V_2 to the same degree and are therefore of no consequence.

Obviously one would think that the l.f. auxiliary signal could be drawn from the a.c. mains supplying the amplifier, but this is undesirable for the following reasons:

Owing to the very low frequency of the mains large capacitors are necessary to smooth out the voltages V_1 and V_2 sufficiently, and this causes

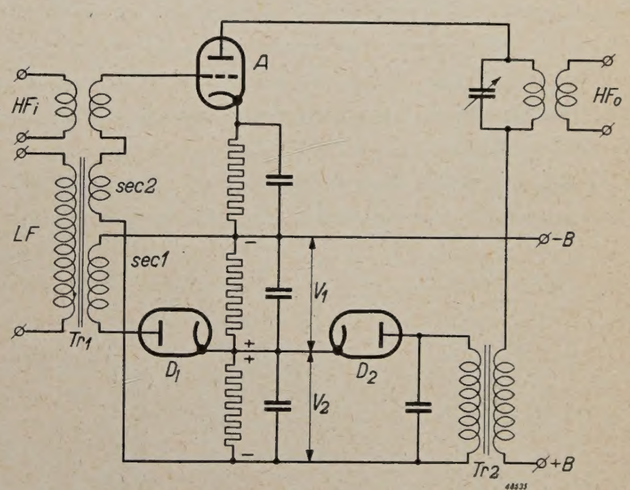


Fig. 5. Amplifier for high frequency (input HF_i, output HF_o) to which a l.f. auxiliary signal *LF* is applied together with the signal to be amplified. The winding *sec 1* on the transformer Tr_1 supplies a voltage V_1 which, after being rectified by the diode D_1 acts positively upon the control grid of the amplifying valve *A*. The voltage from winding *sec 2* is amplified together with the main signal and *via* the transformer Tr_2 (tuned to the low frequency) and the diode D_2 gives a rectified voltage V_2, which just neutralizes V_1 when the amplification is of the nominal value. In case of any deviation from the nominal value then the difference between V_1 and V_2 acts as a compensating control voltage. $+B$ and $-B$ are poles of the anode voltage source.

such a lag in the working that the control cannot respond quickly enough to sudden fluctuations in the mains voltage, with the result that there may be temporarily considerable deviations from the nominal amplification. For that reason it is preferable to employ an auxiliary signal with a higher frequency, say of the order of 1000 c/sec. The drawback of having to generate this signal separately is overcome by the following method, where the amplifier itself produces the auxiliary voltage.

IIb) Controlling the slope with an auxiliary voltage generated in the amplifier

From the circuiting system of fig. 5 it is only a short step to the much better solution of causing the amplifier itself to oscillate in the desired non-interfering auxiliary frequency. For this purpose a signal has to be drawn from the output side and fed back to the input *via* a transfer impedance, in the manner described under Ia) except that whereas with that method the signal fed back has the frequency of the signal to be amplified and it may on no account be allowed to oscillate, in this case the amplifier does oscillate, preferably with a frequency differing from that of the signal to be amplified. In this state of oscillation we have

$$SZ = -1, \quad \ldots \ldots \ldots \quad (8)$$

so that the slope S is fixed, since it must equal a given admittance $-1/Z$. If, therefore, just this admittance has been chosen equal to the desired value of the slope, the occurrence of oscillation proves that the slope is indeed of that value. Deviations from that value are corrected by a control voltage.

Whereas with the method IIa) this control voltage had to be taken from the amplified auxiliary voltage by equation with a fixed voltage, here the rectified oscillation voltage itself is used as control voltage. Thanks to this, smaller auxiliary voltages can be employed, so that there will be much less modulation of the auxiliary signal upon the h.f. signal, which in the circuiting of fig. 5 always occurs more or less. The circuiting, too, can be much simpler, as will be evident presently.

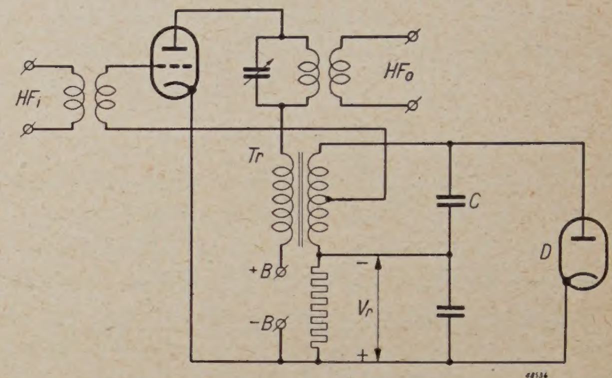


Fig. 6. H.f. amplifier (input HF_i, output HF_o) analogous to that of fig. 5 but with the l.f. auxiliary signal generated in the amplifier itself (oscillating circuit with feedback, consisting of transformer Tr and capacitor C). The l.f. voltage is rectified by a diode D and forms directly the control voltage V_2, used for compensating deviations in slope. $+B$ and $-B$ are poles of the anode voltage source.

The question may now arise whether an exactly constant amplification is really attained in this way. This would indeed be the case if S were not dependent upon the signal amplitude, or, in other words, if the valve characteristic were perfectly linear in the working zone. Actually the valve characteristic is more or less curved; the quantity S in eq. (8) is a sort of average slope of the part of the characteristic traversed, and therefore more or less dependent upon the amplitude. To this extent the ampli-

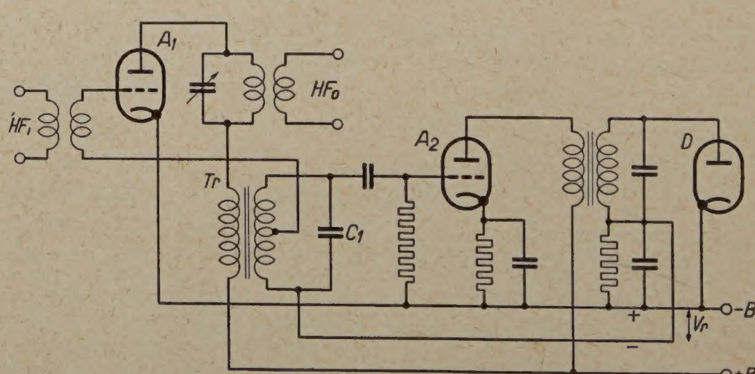


Fig. 7. A variation of the system of fig. 6. The l.f. auxiliary signal generated in the oscillating circuit Tr-C_1 is amplified by the auxiliary valve A_2 before being rectified by the diode D into the control voltage V_r. The advantage here is that the amplitude of the voltage across the circuit Tr-C_1 and also the fraction of it fed to the control grid of the main valve A_1 can be kept very low, with a correspondingly reduced chance of modulation of the l.f. signal upon the h.f. one. $+B$ and $-B$ are poles of the anode voltage source.

fication is not absolutely constant, but with small amplitude the deviations will be only very small.

In the case where the characteristic is absolutely or practically straight the control voltage on the control grid has of course little or no effect, for the slope is (practically) independent of the bias on the control grid. The slope can be influenced, however, by varying a d.c. voltage on the third grid (suppressor grid) of the valve, which must then be a pentode. In such a case, therefore, the control voltage should be applied to the suppressor grid.

We will now give some examples of systems where this method is applied. In *fig. 6* HF_i and HF_0 again represent respectively the input and output terminals of the h.f. signal. The primary coil of the transformer *Tr* is taken up in the anode circuit of an amplifying valve, while the secondary coil forms a l.f. oscillating circuit with the capacitor *C*; part of the a.c. voltage across this circuit is fed back to the control grid, bringing about the required feedback. The l.f. voltage is rectified by the diode *D* to the control d.c. voltage V_r, which, acting negatively upon the control grid, exercises a correcting influence. A comparison of fig. 6 with fig. 5 shows that the former is much simpler.

If it is desired to keep the auxiliary a.c. voltage in the amplifying valve exceptionally low, in order to reduce still further its modulation on the h.f. signal, then the oscillation voltage can be amplified with a separate valve before drawing the control voltage from it. *Fig. 7* gives an example of this, explained in the text underneath.

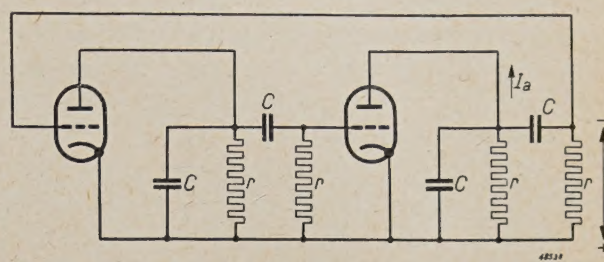


Fig. 8. Example of a system (see footnote [2]) without oscillating circuit, which can still oscillate (so-called *RC* generator).

As already mentioned earlier on, to cause a circuit to oscillate in a certain frequency it is necessary to feed back to the input a voltage with the right phase and amplitude. It is not at all necessary to do this *via* oscillation circuits, for any suitably chosen feedback system consisting only of resistances and capacitors or of resistances and self-inductions (*RC* or *RL* systems) can serve equally well. Especially with low frequencies,

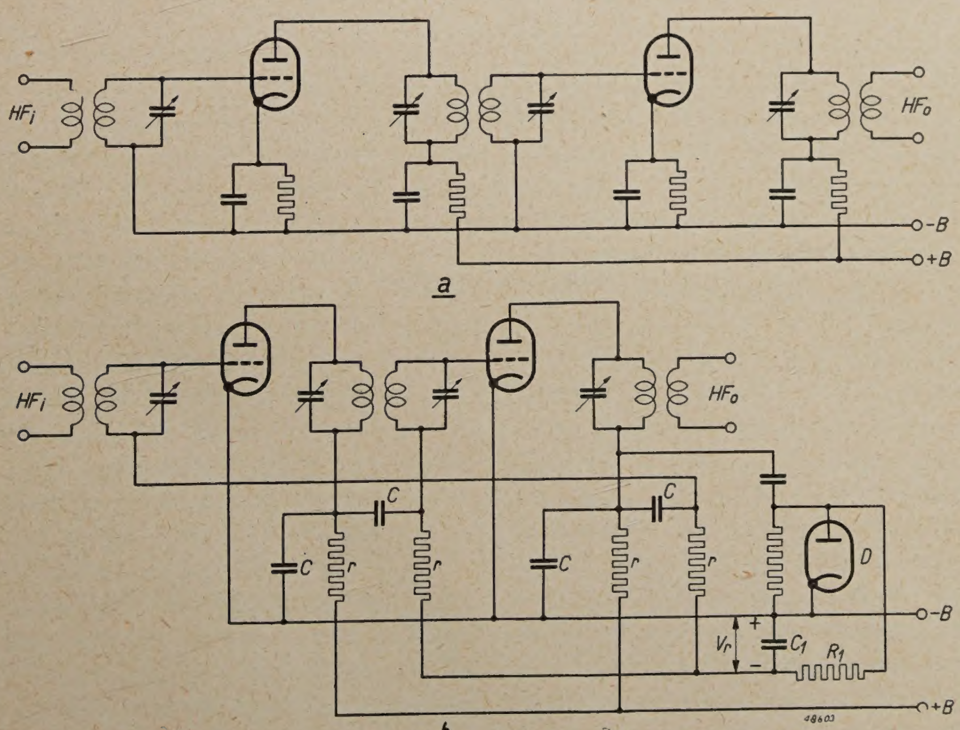


Fig. 9. H.f. amplifier in two stages, *a*) without and *b*) with stabilized amplification. In the latter case the amplifier acts as an *RC* generator on the principle of fig. 8 and thus produces itself the l.f. auxiliary voltage required. By means of the diode *D* and the smoothing circuit R_1-C_1 a negative grid bias V_2 is derived which not only acts as control voltage but also has the function of amplitude limiter. HF_i = input, HF_o = output of the h.f. signal. $+B$ and $-B$ are poles of the anode voltage source. The resistances and the capacitors denoted by *r* and *C* form *RC* systems analogous to that of fig. 8.

where oscillating circuits would involve large and expensive coils, an RC system may be much more economical. Some of these circuiting systems are already familiar as generators of voltages with rectangular or other non-sinusoidal curves, such as Abraham and Bloch's multivibrator. It is less known, however, that if in these circuits only the amplitude is limited in the right way the oscillations remain practically sinusoidal. They are then very usuful for stabilizing the slope of an amplifier in the manner just described.

Of the many circuiting systems that might usefully be employed *fig. 8* gives a simple example (see footnote [2]). The only difference from fig. 2 is that the resistance and capacitor farthest to the left in fig. 2 are changed round and two stages are in cascade connection (with only one stage the voltage V_t would have to be displaced 180° in phase to allow of oscillation). In fig. 8 we get for the transfer impedance $Z = V_t/I_a$:

$$Z = \frac{-3\,p^2 + jp\,(1-p^2)}{1 + 7\,p^2 + p^4} \cdot r, \quad . \quad . \quad . \quad (9)$$

where p again represents ωCr. As will be seen, Z can now be made a real value for $p = 1$, for which it assumes the value $-r/3$. Oscillation will therefore occur if $S = 3/r$, and then with the frequency $f = 1/2\pi Cr$.

Inversely, when the circuit does actually oscillate one can conclude that the slope is $S = 3/r$ and is therefore fixed by the resistance value r.

Limitation of the amplitude can be effected in a simple manner by rectifying and smoothing the a.c. voltage generated and using the resultant d.c. voltage (or a part of it) as grid bias for the oscillating amplifier valve. One finds this applied in the two-stage amplifier for high frequency shown in *fig. 9b* (on the principle of fig. 8). A comparison with fig. 9*a* of the same amplifier without stabilized amplification shows that only a few, inexpensive components are required for stabilizing, and there is no loss of amplifying power.

In conclusion it is to be observed that two or more of the methods described here can be combined for raising the degree of stability in amplification to an exceptionally high level.